TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+, NH_4^+, P_i, SO_4^{2-}, sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging techniques

and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., bi tin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H^+-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H^+-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na^+-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective

transporters for organic cations and organic anions in organs including the kidn y, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) Am. J. Physiol. 264:C761-C782; Price, N.T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and I. Haggstrom (1993) J. Biotechnol. 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) Meth. Enzymol. 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process.

Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

Ion Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the

energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient. These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na^+-K^+ ATPase, Ca^{2+}-ATPase, and H^+-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na^+ and Ca^{2+} are low and cytosolic concentration of K^+ is high. The vacuolar (V) class of ion transporters includes H^+ pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H^+ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the V_1 domain, a peripheral complex responsible for ATP hydrolysis; and the V_0 domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F_0 domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V_0 domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na^+ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca^{2+} out of the cell with transport of Na^+ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated chann ls open their pores in

response to mechanical stress; voltage-gated channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca^{2+} and Na^+ (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na^+ and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^+ channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^+ channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^+ and K^+ channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^+ and K^+ ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^+ channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na^+ channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^+ channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, β1 and β2. The β2 subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α and β1 subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H^+-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from C. elegans. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitable tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+-K^+ pump and ion channels that provide the redistribution of Na^+, K^+, and Cl^-. The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form

functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated K^+ channels as well as the d layed rectifier type channels such as the human ether-a-go-go related gene (HERG) associated with long QT, a cardiac dysrythmia syndrome (Curran, M.E. (1998) Curr. Opin. Biotechnol. 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) Curr. Opin. Chem. Biol. 3:448-458).

A second superfamily of K^+ channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting K^+ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K^+ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) Curr. Opin. Neurobiol. 5:268-277; Curran, supra).

The recently recognized TWIK K^+ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) EMBO J 16:5464-5471).

The voltage-gated Ca^{2+} channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca^{2+} channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca^{2+} channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the $\alpha_2\delta$ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six α_1, one $\alpha_2\delta$, and four β genes. A fourth subunit, γ, has been identified in skeletal muscle (Walker, D. et al. (1998) J. Biol. Chem. 273:2361-2367; McCleskey, E.W. (1994) Curr. Opin. Neurobiol. 4:304-312).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca^{2+} influx into cells to resupply Ca^{2+} stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from Drosophila and have similarity to voltage

gated Ca2+ channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCC entry channels (Zhu, X. et al. (1996) Cell 85:661-671; Boulay, G. et al. (1997) J. Biol. Chem. 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, and whose expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo. The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) J. Clin. Oncol. 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+, K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the voltage-gated channels and starts th action potential. Chloride channels open in response to inhibitory neurotransmitters,

such as γ-aminobutyric acid (GABA) and glycine, leading to hyperpolarization of th membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, supra). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol. 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the Gβγ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate

kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) Cell 93:495-498).

Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, <u>supra</u>).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na^+ channels have been useful in the treatment of neuropathic pain (Eglen, <u>supra</u>).

Ion channels in the immune system have recently been suggested as targets for

immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2,""TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," "TRICH-15," "TRICH-16," "TRICH-17," "TRICH-18," "TRICH-19," "TRICH-20," "TRICH-21," "TRICH-22," "TRICH-23," "TRICH-24," "TRICH-25," and "TRICH-26." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In ne alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-26. In

another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:27-52.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to

the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The invention additionally

provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotid selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a

polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotid complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH.

The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. D liberate amino acid substitutions may be mad on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies

which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to licit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which act on right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical

functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu

Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides r amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:27-52 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:27-52, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:27-52 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:27-52 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:27-52 and the region of SEQ ID NO:27-52 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-26 is encoded by a fragment of SEQ ID NO:27-52. A fragment of SEQ ID NO:1-26 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-26. For example, a fragment of SEQ ID NO:1-26 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-26. The precise length of a fragment of SEQ ID NO:1-26 and the region of SEQ ID NO:1-26 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e

sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a

length over which percentage identity may b measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be

used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) <u>Molecular Cloning: A Laboratory Manual</u>, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances,

such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisens molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such

purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, <u>supra</u>. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions

(UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissu under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient

cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or <u>in vitro</u> fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), <u>supra</u>.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene

between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological, and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte

p lypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:2 is 94% identical from amino acids 965 through 2436 to mouse abc2 transporter (GenBank ID g495259) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains two ABC transporter domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is an ABC transporter. In an alternate example, SEQ ID NO:13 is 97% identical to human gamma subunit precursor of muscle acetylcholine receptor (GenBank ID g825618) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.0e-273, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:13 also contains a neurotransmitter-gated ion-channel domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is a neurotransmitter-gated ion-channel protein. In an alternate example, SEQ ID NO:19 is 62% identical to human vacuolar proton-ATPase (GenBank ID g37643) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.2e-129, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:19 is a vacuolar ATP synthase. In an alternate exampple, SEQ ID NO:22 is 94% identical to rat GABA(A) receptor gamma-1 subunit (GenBank ID g56176) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.4e-244, which indicates the probability f obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also contains a neurotransmitter-gated ion channel domain as determined by searching for statistically significant

matches in the hidden Markov model (HMM)-based PFAM databas of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:22 is a neurotransmitter-gated ion channel. In an alternate example, SEQ ID NO:26 is 61% identical to rabbit peroxisomal Ca-dependent solute carrier (GenBank ID g2352427) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.4e-156, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:26 also contains three mitochondrial carrier protein domains, as well as three EF hand domains, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:26 is a calcium dependent carrier protein. In an alternate examplpe, SEQ ID NO:17 is 69% identical to *Ambystoma tigrinum* electrogenic $NaHCO_3$ cotransporter (GenBank ID g2198815) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains an HCO_3 transporter family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is an anion transporter. SEQ ID NO:1, SEQ ID NO:3-12, SEQ ID NO:14-16, SEQ ID NO:18, and SEQ ID NO:20-25 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-26 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:27-52 or that distinguish between SEQ ID NO:27-52 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5')

and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7251266F7 is the identification number of an Incyte cDNA sequence, and PROSTMY01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70564238V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g4689801) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_*gAAAAA*_*gBBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-52, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:27-52, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-

52 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:27-52. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:27-52 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment

of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which ften include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These

preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutati ns may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. How ver, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided

by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of

transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers,

or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These

procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation

lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a

natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science

244:1288-1292). The vector integrates into the corresponding regi n of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, the expression of TRICH is closely associated with brain, lung, prostate, bladder, bone, hypothalamus, breast, ileum, stomach, pancreas, and gastrointestinal tissues and tumors of the brain and prostrate. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological, and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder

such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder

such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystr phy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus; systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for

use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and

ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translati n f mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavag . For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of

polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved

using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic

purposes may b prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:27-52 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA

polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal dis rders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular

disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may b used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in

dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any f the polynucleotide sequences described herein may be used as elements on a microarray. The microarray

can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression <u>in vivo</u>, as in the case of a tissue or biopsy sample, or <u>in vitro</u>, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with <u>in vitro</u> model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a

signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or

untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared

with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to th test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery

techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below and including U.S. Ser. No. 60/232,685, U.S. Ser. No. 60/234,842, U.S. Ser. No. 60/236,882, U.S. Ser.

No. 60/239,057, U.S. Ser. No. 60/240,540, and U.S. Ser. No. 60/241,700 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA (Invitrogen), PCMV-ICIS (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family

databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:27-52. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is

a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

<u>"Stitched" Sequences</u>

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear

along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:27-52 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:27-52 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in

humans, although this can vary widely due to hot and cold spots of recombination.) Th cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were includ d in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:31 was mapped to chromosome 1 within the interval from 133.00 to 137.30 centiMorgans. SEQ ID NO:33 was mapped to chromosome 12 within the interval from 120.50 to the q terminal, or more specifically, within the interval from 126.10 to 145.70 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the

other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min;

Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:27-52 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of olig nucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the

biological sampl are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

<u>**Hybridization**</u>

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

<u>**Detection**</u>

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is

typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophag promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII where applicable.

XIII. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a

marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the ligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

XVI. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as $G\beta\gamma$ proteins (Reimann, supra) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, supra). TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

XVII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18°C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of th TRICH mediator, such as cAMP, cGMP, or Ca^{+2} (in the form of $CaCl_2$), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking

mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to th activity of TRICH in the assay. In particular, the activity of TRICH-25 is measured as Cl- conductance.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH, 50μg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^{3}H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate. In particular, test substrates include amino acids for TRICH-1, xanthine and uracil for TRICH-3, melibiose for TRICH-18, monocarboxylate for TRICH-20, neurotransmitters such as gamma-aminobutyric acid (GABA) for TRICH-22, and nucleosides for TRICH-23.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-^{32}P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP-[γ-^{32}P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

XVIII. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal i n concentration are measured using fluorescent dyes such as the

Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^- indicator MQAE (all available fr m Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as $DiBAC_4$ (Molecular Probes). $DiBAC_4$ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of $DiBAC_4$ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1687189	1	1687189CD1	27	1687189CB1
7078207	2	7078207CD1	28	7078207CB1
1560619	3	1560619CD1	29	1560619CB1
2614283	4	2614283CD1	30	2614283CB1
2667691	5	2667691CD1	31	2667691CB1
3211415	6	3211415CD1	32	3211415CB1
4739923	7	4739923CD1	33	4739923CB1
55030459	8	55030459CD1	34	55030459CB1
6113039	9	6113039CD1	35	6113039CB1
7101781	10	7101781CD1	36	7101781CB1
7473036	11	7473036CD1	37	7473036CB1
7476943	12	7476943CD1	38	7476943CB1
8003355	13	8003355CD1	39	8003355CB1
3116448	14	3116448CD1	40	3116448CB1
622868	15	622868CD1	41	622868CB1
7476494	16	7476494CD1	42	7476494CB1
7477260	17	7477260CD1	43	7477260CB1
1963058	18	1963058CD1	44	1963058CB1
2395967	19	2395967CD1	45	2395967CB1
3586648	20	3586648CD1	46	3586648CB1
7473396	21	7473396CD1	47	7473396CB1
7476283	22	7476283CD1	48	7476283CB1
7477105	23	7477105CD1	49	7477105CB1
7482079	24	7482079CD1	50	7482079CB1
55145506	25	55145506CD1	51	55145506CB1
5950519	26	5950519CD1	52	5950519CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	1687189CD1	g2116552	7.80E-275	[Rattus norvegicus] cationic amino acid transporter 3 (Hosokawa,H. et al. (1997) J. Biol. Chem. 272 (13), 8717-8722)
2	7078207CD1	g495259	0	[Mus musculus] abc2 (Luciani,M.F. et al. (1994) Genomics 21 (1), 150-159)
3	1560619CD1	g1002424	8.60E-253	[Mus musculus] YSPL-1 form 1 (Guimaraes,M.J. et al. (1995) Development 121 (10), 3335-3346)
4	2614283CD1	g1256378	1.90E-152	[Rattus norvegicus] zinc transporter ZnT-2 (Palmiter,R.D. et al. (1996) EMBO J. 15 (8), 1784-1791)
5	2667691CD1	g2506078	2.30E-259	[Mus musculus] tetracycline transporter-like protein (Matsuo,N. et al. (1997) Biochem. Biophys. Res. Commun. 238 (1), 126-129)
6	3211415CD1	g7243710	9.80E-197	[Mus musculus] zinc transporter like 2
7	4739923CD1	g13785620	4.00E-96	[3' incom][Mus musculus] sideroflexin 5 (Fleming,M.D. et al. (2001) Genes Dev. 15 (6), 652-657)
8	55030459CD1	g4186073	9.40E-15	[Mus musculus] calcium channel alpha-2-delta-C subunit (Klugbauer,N. et al. (1999) J. Neurosci. 19(2), 684-691)
9	6113039CD1	g310183	5.00E-273	[Rattus norvegicus] sodium dependent sulfate transporter (Markovich,D. et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 8073-8077)
10	7101781CD1	g13506808	0	[fl][Mus musculus] thymic stromal co-transporter (Chen,C. et al. (2000) Biochim. Biophys. Acta 1493 (1-2), 159-169)
11	7473036CD1	g13249295	0	[fl][Homo sapiens] anion exchanger AE4 (Parker,M.D. et al. (2001) Biochem. Biophys. Res. Commun. 282 (5), 1103-1109)
12	7476943CD1	g3047402	5.00E-67	[Homo sapiens] monocarboxylate transporter 2
13	8003355CD1	g825618	3.00E-273	[Homo sapiens] ach_cds (Shibahara,S. et al. (1985) Eur. J. Biochem. 146 (1), 15-22)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
14	3116448CD1	g10732815	0	[fl][Homo sapiens] concentrative Na+-nucleoside cotransporter hCNT3 (Ritzel,M.W.L. et al. (2001) J. Biol. Chem. 276 (4), 2914-2927)
15	622868CD1	g5924012	9.50E-160	[Homo sapiens] dJ261K5.1 (novel organic cation transporter (BAC ORF RG331P03))
16	7476494CD1	g8979801	3.90E-147	[Homo sapiens] dJ37C10.3 (novel ATPase)
17	7477260CD1	g13447747	0	[fl][Homo sapiens] sodium bicarbonate cotransporter NBC4a (Pushkin,A. et al. (2000) IUBMB Life 50 (1), 13-19)
18	1963058CD1	g1653342	6.80E-18	[Synechocystis sp.] melibiose carrier protein (Kaneko,T. et al. (1995) DNA Res. 2 (4), 153-166)
19	2395967CD1	g37643	3.20E-129	[Homo sapiens] vacuolar proton-ATPase (van Hille,B. et al. (1993) Biochem. Biophys. Res. Commun. 197 (1), 15-21)
20	3586648CD1	g2198807	8.90E-49	[Gallus gallus] monocarboxylate transporter 3
		g2463628	6.00E-43	[fl][Homo sapiens] putative monocarboxylate transporter
21	7473396CD1	g2618842	1.10E-139	[Bacillus subtilis] excinuclease ABC subunit (Reizer,J. et al. (1998) Mol. Microbiol. 27 (6), 1157-1169)A
22	7476283CD1	g56176	4.40E-244	[Rattus norvegicus] GABA(A) receptor gamma-1 subunit (Ymer,S. et al. (1990) EMBO J. 9 (10), 3261-3267)
23	7477105CD1	g3176684	2.20E-11	[Arabidopsis thaliana] Contains similarity to equilibratiave nucleoside transporter 1 gb|U81375 from Homo sapiens. ESTs gb|N65317, gb|T20785, gb|AA586285 and gb|AA712578 come from this gene
		g12656639	3.00E-05	[fl][Homo sapiens] equilibrative nucleoside transporter 3
24	7482079CD1	g2815899	9.60E-84	[Homo sapiens] Shab-related delayed-rectifier K+ channel alpha (Shepard,A.R. et al. (1999) Am. J. Physiol. 277 (3), C412-C424)
25	55145506CD1	g289404	4.70E-105	[Bos taurus] chloride channel protein (Landry,D. et al. (1993) J. Biol. Chem. 268, 14948-14955)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
26	5950519CD1	g2352427	6.40E-156	[Oryctolagus cuniculus] peroxisomal Ca-dependent solute carrier (Weber,F.E. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94 (16), 8509-8514)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1687189CD1	619	S134 S33 S453 S589 S599 T104 T18 T220 T272 T273 T438 T451 Y224	N232	Transmembrane domains: C31-Y51, S65-A85, D165-A183, V196-V214, P383-F401, M410-L428, V479-W498, L508-W528, A543-M562, W567-I593	HMMER
					Amino acid permeases signature BL00218: I66-A97, C343-T382	BLIMPS_BLOCKS
					CATIONIC AMINO ACID TRANSPORTER PD034711: Q431-I523	BLAST_PRODOM
					AMINO ACID CATIONIC TRANSPORTER TRANSPORT TRANSMEMBRANE GLYCOPROTEIN TRANSPORTER1 PROTEIN HIGHAFFINITY PD000262: V526-Q597	BLAST_PRODOM
					TRANSMEMBRANE TRANSPORT PROTEIN TRANSPORTER AMINOACID PERMEASE AMINO ACID GLYCOPROTEIN MEMBRANE PD000214: L28-L428	BLAST_PRODOM
					do ANTIPORTER; ORNITHINE; PUTRESCINE; TRANSPORT; DM01125|P30825|23-373: E25-R371	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	7078207CD1	2436	S1114 S1119 S1133 S119 S1248 S1323 S1332 S1339 S1381 S140 S1411 S1427 S1455 S1478 S1560 S1604 S1687 S1819 S1982 S199 S2024 S2036 S2062 S21 S2159 S2196 S2245 S2292 S2333 S2366 S2420 S256 S281 S467 S50 S502 S533 S631 S884 S940 S959 S971 T1058 T1081 T1212 T1271 T1313 T1314 T1532 T16 T2097 T2102 T2108 T2144 T2215 T2235 T2284 T2352 T2413 T252 T353 T382 T440 T48 T612 T633 T696 T844 T955 Y1390	N14 N1409 N1497 N1550 N1558 N1613 N1678 N169 N174 N1776 N2055 N306 N369 N380 N421 N433 N477 N485 N495 N531 N545 N591 N601 N629 N90	Transmembrane domains: P22-K45, V784-L803, L893-T911, V1793-F1813, M1845-F1862, V1900-L1926	HMMER
					ABC transporter domains: N1018-G1198, G2081-G2262	HMMER_PFAM
					ABC transporters family signature: D1105-D1155, V2167-D2218	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2					ATPBINDING TRANSPORTER CASSETTE ABC TRANSPORT PROTEIN GLYCOPROTEIN TRANSMEMBRANE RIM ABCR PD005939: L1787-Y1971 PD006867: I663-S809 PD006285: A811-K1001 PD010138: G1957-K2061	BLAST_PRODOM
					ABC TRANSPORTERS FAMILY DM00008|P41233|839-1045: V991-H1197, V2051-M2259	BLAST_DOMO
					ABC transporter motif: L1124-F1138	MOTIFS
					ATP/GTP binding site (P-loop): G1025-T1032, G2088-T2095	MOTIFS
					Lipocalin motif: G1424-V1437, G1426-V1437	MOTIFS
3	1560619CD1	610	S127 S161 S251 S409 S450 S483 S582 S601 S608 T313 T514 T529	N139 N159	Transmembrane domains: F215-C233, L264-I286	HMMER
					Xanthine/uracil permeases family domain: G46-E473	HMMER_PFAM
					Xanthine/uracil permease signature BL01116: G407-F443	BLIMPS_BLOCKS
					YOLK SAC PERMEASELIKE YSPL1 FORM 1 YOLK SAC PERMEASELIKE YSPL1 FORM 4 YOLK SAC PERMEASELIKE YSPL1 FORM 3 YOLK SAC PERMEASELIKE YSPL1 FORM 2 PD019501: G429-Q609 PD137940: Q29-P83	BLAST_PRODOM
					PROTEIN TRANSPORT SULFATE TRANSPORTER TRANSMEMBRANE PERMEASE INTERGENIC REGION AFFINITY GLYCOPROT PD001255: L174-L467	BLAST_PRODOM
					XANTHINE/URACIL PERMEASES FAMILY DM01485|S33349|7-188: G355-L465	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	2614283CD1	372	S124 S216 S338 S61 T281		Transmembrane domain: I141-V159	HMMER
					Cation efflux family: P127-S358	HMMER_PFAM
					ZINC TRANSPORTER CATION EFFLUX COBALT RESISTANCE PD001369: N214-S371 PD001602: Q71-H197	BLAST_PRODOM
					ZINC TRANSPORTER ZNT2 PD095371: A15-C81	BLAST_PRODOM
					TRANSPORTER; EFFLUX; ZINC; CZCD DM02892|P13512|9-157: G68-S204 DM02892|P20107|1-136: R72-T207 DM02892|P32798|2-127: R72-S168 DM02892|S54302|3-128: G68-I191	BLAST_DOMO
5	2667691CD1	490	S212 S236 S455 S460 T406	N12 N453	Transmembrane domains: A40-V61, P123-V147, V191-V209, D243-Y257, V282-S302, L432-I448	HMMER
					TETRACYCLINE RESISTANCE PR01035: I130-T151, Y160-G182, P429-P449, V282-S302, W335-S355, V370-F393	BLIMPS_PRINTS
					HIPPOCAMPUS ABUNDANT PROTEIN TRANSCRIPT 1 TETRACYCLINE TRANSPORTER LIKE PROTEIN PD125679: Y394-V490 PD082602: M1-H39	BLAST_PRODOM
					Sugar transport proteins signatures: I92-T108	MOTIFS
6	3211415CD1	377	S223 S31 S321 S5 T338 T34 Y98	N45	Transmembrane domains: L106-S124, R140-F159, I270-L287	HMMER
					Cation efflux family: R91-A376	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6					ZINC TRANSPORTER CATION EFFLUX COBALT RESISTANCE PD001602:L38-V158	BLAST_PRODOM
					TRANSPORTER; EFFLUX; ZINC; CZCD DM02892|S61568|396-545: D32-H166 DM02892|P20107|1-136: I29-S170 DM02892|P13512|9-157: R36-G162 DM02892|P32798|2-127: L42-I154	BLAST_DOMO
7	4739923CD1	340	S240 S272 S314 S319 S330 T56 T74	N127 N140 N153	CHROMOSOME PUTATIVE TRANSPORTER C17G6.15C TRANSPORT XV READING FRAME PD006986: S20-L274	BLAST_PRODOM
8	55030459CD1	1274	S1025 S1138 S1142 S1155 S1189 S1201 S1242 S134 S190 S238 S256 S298 S303 S353 S354 S40 S405 S430 S624 S664 S670 S700 S746 S79 S892 S894 S952 T1006 T1050 T1191 T221 T268 T272 T293 T349 T361 T581 T674 T717 T75 T755 T813 T852 T868 T987 Y1056 Y114	N145 N329 N373 N568 N587 N905 N940 N985	signal_cleavage: M1-A35	SPSCAN
					Transmembrane domain: V1096-R1118	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	6113039CD1	595	S213 S214 S483 S74 T209 T230 T236 T240 T423 T97 Y39	N140 N174 N207 N591	Transmembrane domains: Y10-L30, F287-W305, V349-D365, G556-M576	HMMER
					Sodium:sulfate symporter BL01271: T131-I150, T240-I264, P432-G453, A505-I559	BLIMPS_BLOCKS
					SODIUM SYMPORT OF COTRANSPORTER PD000549: E331-W572, L242-K402, V16-A167, F13-V154	BLAST_PRODOM
					SODIUM/SULFATE COTRANSPORTER NA+/SULFATE TRANSPORT TRANSMEMBRANE SODIUM SYMPORT PD084897: A161-K238	BLAST_PRODOM
					do RENAL; BOUND; PRO-SER-ALA; NA; DM02914|A47714|28-576: I28-F577 DM02914|S43561|28-507: L242-I569, E34-A161 DM02914|P46556|1-520: K198-F577, E34-I159 DM02914|P32739|25-517: K238-F577, E34-V154	BLAST_DOMO
10	7101781CD1	475	S100 S108 S170 S34 S61 T20 T252 T390	N55	Transmembrane domains: I283-V308, Y322-V340, M350-E370, S440-V459	HMMER
11	7473036CD1	927	S149 S163 S217 S23 S260 S265 S325 S51 S65 S733 S738 S874 S891 S904 S906 T292 T324 T344 T567 T594 T629 T802 T99	N493 N520 N544 N923	Transmembrane domains: V444-Y466, V761-P780, I779-M810	HMMER

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>11</td><td></td><td></td><td></td><td></td><td>HCO3- transporter family:
K108-I835</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Anion exchangers family
BL00219: V360-D383, W659-L700, G744-L789, Y790-T833, G89-H120, Q180-L223</td><td>BLIMPS_BLOCKS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Anion exchangers family signatures:
A457-G509</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ANION EXCHANGER SIGNATURE
PR00165: Q355-G375, V388-G407, L442-S461, G474-L492, D570-L589, W657-M676</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ANION EXCHANGE GLYCOPROTEIN PALMITATE BICARBONATE COTRANSPORTER
PD001455: S346-L784, S505-I835, S156-F348, L109-V154</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>BICARBONATE COTRANSPORTER ELECTRO-GENIC NA+ PANCREAS HCO3 F52B5.1
PD018437: Q836-N927</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>BAND 3 ANION TRANSPORT PROTEIN
DM02294|P04920|602-1237: G558-E894, S346-P529
DM02294|P48751|601-1229: S537-G896, S346-I543
DM02294|A42497|403-1027: S537-G896, S346-I500
DM02294|P02730|311-908: P560-D882, S346-G519</td><td>BLAST_DOMO</td></tr>
<tr><td>12</td><td>7476943CD1</td><td>516</td><td>S11 S137 S169 S202 S253 S41 S92 T228 T234 T244 T30 T340</td><td>N10 N333 N487</td><td>Transmembrane domains:
I118-T144, S181-W203, A206-M224, Y275-M293</td><td>HMMER</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12					Monocarboxylate transporter: C55-D499	HMMER_PFAM
					PEST; TRANSPORTER; LINKED; DM05037|P53988|1-465: P42-Q470 DM05037|Q03064|1-475: S41-D479 DM05037|P36021|155-612: A37-L258, V285-E477	BLAST_DOMO
13	8003355CD1	514	S174 S183 S330 S427 S453 S54 S64 T381 T382 Y94	N163 N328 N373 N52	signal_cleavage: M1-G22	SPSCAN
					signal peptide: M1-G22	HMMER
					Transmembrane domains: P241-F264, C274-V291, Y308-N328, V472-M491	HMMER
					Neurotransmitter-gated ion-channel: E26-F489	HMMER_PFAM
					Neurotransmitter-gated ion-channels proteins BL00236: V107-N116, D135-Y173, H228-A269, V53-D90	BLIMPS_BLOCKS
					Neurotransmitter-gated ion-channels signature: V130-Q184	PROFILESCAN
					Neurotransmitter-gated ion channel family signature PR00252: T73-R89, M106-N117, C150-C164, L235-N247	BLIMPS_PRINTS
					Nicotinic acetylcholine receptor sig. PR00254: T60-V76, Y94-W108, I112-V124, V130-S148	BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					CHANNEL IONIC GLYCOPROTEIN POSTSYNAPTIC RECEPTOR SIGNAL PROTEIN PD000153: N24-S393, A432-F489	BLAST_PRODOM
					NEUROTRANSMITTER-GATED ION-CHANNELS DM00195|P13536|7-501: P7-V497 DM00195|P02713|5-498: L8-R496 DM00195|P05376|2-493: L10-R496 DM00195|P02714|1-491: L8-V497	BLAST_DOMO
					Neurotransmitter-gated ion-channels signature: C150-C164	MOTIFS
14	3116448CD1	691	S326 S36 S549 S582 S63 S669 T100 T193 T262 T356 T411 T417 T50 T615 T637 Y87	N30 N34 N630 N636 N664	Transmembrane domain: I104-N124, W178-L207, L289-M308, I444-L461	HMMER
					Na+ dependent nucleoside transporter Nucleoside_tra2: Q198-S613	HMMER_PFAM
					Copper-transporting ATPase L131-D145	BLIMPS_PRINTS
					NA+/NUCLEOSIDE INNER MEMBRANE TRANSPORT PD003768: R223-I611, PD008773: F93-F215	BLAST_PRODOM
					NUCLEOSIDE; TRANSPORT; NaDEPENDENT DM01857|A54892|234-589: L256-L612 DM01857|A57532|230-585: L256-L612 DM01857|P44742|60-409: V260-I611 DM01857|P33021|60-412: V260-G610	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	622868CD1	342	S102 S309 S315 S325 S84 T121 T174 T286 T299 T300 T334	N110 N117 N311 N323	Transmembrane domain: Y205-Y227	HMMER
					PERIPHERIN (RDS)/ROM-1 F PR00218: V9-V29, L207-L228	BLIMPS_PRINTS
					SUGAR TRANSPORTER SIGNATURE PR00171: A231-V242	BLIMPS_PRINTS
					DM00135|P39932|141-478: W33-K295	BLAST_DOMO
16	7476494CD1	791	S103 S110 S199 S289 S514 S659 S66 S688 S734 S782 T122 T191 T287 T314 T326 T507 T710 T747 T78 Y293 Y742	N697 N768	Atpase_E1_E2 D437-T443	MOTIFS
					transmembrane domain: A177-Y193, D348-Y366	HMMER
					E1-E2 ATPase E1-E2_ATPase: C217-T443, P551-R680	HMMER_PFAM
					E1-E2 ATPases phosphorylation site atpase_e1_e2.prf: I417-A471	PROFILESCAN
					E1-E2 ATPases phosphorylation site BL00154: V393-G429, L431-L449, K575-C585, N644-M684	BLIMPS_BLOCKS
					P-type cation-transporting ATPase superfamily signature PR00119: D260-T274, C435-L449, A660-D670	BLIMPS_PRINTS
					Sodium/potassium-transporting ATPase signature PR00121: C428-L449, A572-V590	BLIMPS_PRINTS
					E1-E2 ATPASES PHOSPHORYLATION SITE DM00115|P22189|49-801: L547-V685 DM00115|P37278|58-755: Q224-I692 DM00115|A42764|65-737: E141-T699 DM00115|P37367|60-746: L226-V691	BLAST_DOMO
					ATPBINDING CALCIUM MAGNESIUM TRANSPORT PUMP PD000132: I180-D445, A612-Q689, M559-C585	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7477260CD1	1108	S1011 S1061 S1063 S1088 S124 S14 S190 S218 S240 S314 S319 S388 S391 S434 S435 S686 S701 S870 S95 T1030 T1056 T1065 T1093 T1102 T16 T183 T201 T454 T639 T678 T725 T766 T778 T78 Y1090	N399 N653 N658 N668 N676	Gene regulatory motif Leucine_Zippe L125-L146, L677-L698	MOTIFS
					Anion exchangers family signatures anion_exchanger1.prf: D438-F490 anion_exchanger2.prf: A585-T639	PROFILESCAN
					Transmembrane domain: I488-L506, L837-W856, I898-P917, V920-F938, I982-V1002	HMMER
					HCO3- transporter family HCO3_cotransp: K104-V972	HMMER_PFAM
					Anion exchangers family BL00219: H85-H116, K259-V302, T304-K342, A343-K378, G448-A487, I488-D511, L541-Q579, L581-I628, P706-D759, V796-L837, D838-E876, G881-L926, Y927-T970, V972-S1011	BLIMPS_BLOCKS
					ANION EXCHANGER SIGNATURE PR00165: F458-F480, Q483-G503, V516-G535, I539-S558, L570-S589, G602-I620, D707-L726, L742-F762, W794-M813	BLIMPS_PRINTS
					BAND 3 ANION TRANSPORT PROTEIN DM02294|P48751|601-1229: P706-N1020, E449-P634, I353-E367	BLAST_DOMO
					PROTEIN ANION EXCHANGE TRANSMEMBRANE BAND GLYCOPROTEIN LIPOPROTEIN PALMITATE BICARBONATE COTRANSPORTER PD001455: H445-V972, V105-E394	BLAST_PRODOM
					BICARBONATE COTRANSPORTER SODIUM ELECTROGENIC NA+ PANCREAS PD018437: Q973-M1078 PD018439: A53-E103	BLAST_PRODOM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>18</td><td>1963058CD1</td><td>480</td><td>S13 S194 S195 S204 S409 S49 S54 T286</td><td>N178 N219 N292 N341</td><td>Transmembrane domain:
L349-Y369</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Sodium:galactoside symporter family:
DM01084|Q02581|1-462:L17-S195 (P-value = 8.2e-10)</td><td>BLAST-DOMO</td></tr>
<tr><td>19</td><td>2395967CD1</td><td>381</td><td>S119 S170 S202 S211 S269 S327 S349 S378 S74 T102 T144 T147 T164 T246 T26 T328 T62</td><td>N192 N285 N30</td><td>Vacuolar ATPase C subunit
PD014267: E3-D376</td><td>BLAST-PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Vacuolar ATP synthase:
DM04365|P21282|1-381: M1-D381
DM04365|P54648|1-368: E3-L342
DM04365|P31412|1-392: I7-L377</td><td>BLAST-DOMO</td></tr>
<tr><td>20</td><td>3586648CD1</td><td>484</td><td>S236 S4 T21 T258 T290 T3 T312 Y301</td><td>N345 N389</td><td>Transmembrane domains:
F42-W61, V75-I94, F311-Y327, I361-W382, W382-M402</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Monocarboxylate transporter:
S40-L478</td><td>HMMER-PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Transporter
DM05037|P53988|1-465: P16-N217
DM05037|Q03064|1-475: D29-Q263
DM05037|P36021|155-612: D29-L229</td><td>BLAST-DOMO</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7473396CD1	736	S236 S440 S462 S472 S501 S52 S579 S626 T10 T166 T191 T239 T316 T324 T345 T386 T491 T587 T607 T715 T89	N3 N367	Signal peptide: M1-G53	SPScan
					ABC transporter: G24-G210, G429-G700	HMMER-PFAM
					ABC transporter: L396-V410, L625-L639	MOTIFS
					ATP/GTP binding sites: G31-S38, G436-S443	MOTIFS
					ABC transporters family signatures: Q606-H659, L378-D427	Profil Scan
					ABC transporters family BL00211: L396-D427, L29-L40	BLIMPS-BLOCKS
					UVRA protein DM02034|P13567|759-959: F503-G704, I135-D202 DM02034|P07671|708-908: F503-G704, I135-D202 DM02034|S49424|2-201: D504-G704, K110-K198 DM02034|P47660|610-810: F503-G704, I135-D202	BLAST-DOMO
					Excinuclease ABC subunit A PD001646: D504-T624	BLAST-PRODOM
					Excinuclease ABC subunit A PD184930: C538-T713, R133-L297, V434-V502	BLAST-PRODOM
					Excinuclease ABC subunit A PD003881: N447-F503	BLAST-PRODOM
					Ribose/galactose ABC transporter: PD035715: K241-K311, M1-I55	BLAST-PRODOM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>22</td><td>7476283CD1</td><td>465</td><td>S153 S224 S267 S426 T118 T129 T179 T331 T88</td><td>N127 N245 N393 N50</td><td>Signal peptide:
M1-C35</td><td>SPScan</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Transmembrane domain:
M270-I294</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel:
I64-W459</td><td>HMMER-PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion-channels signature:
L168-K222</td><td>ProfileScan</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel:
C188-C202</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel
BL00236: I90-N127, I143-N152, D173-Y211, Y257-A298</td><td>BLIMPS-BLOCKS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel:
PR00252: T110-F126, K142-S153, C188-C202, F264-Q276</td><td>BLIMPS-PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Gamma-aminobutyric acid receptor:
PR00253: F273-W293, V299-A320, M333-L354, Y442-Y462</td><td>BLIMPS-PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Gamma-aminobutyric acid receptor:
PR01079: G62-Q73, D82-I99, F125-N138, W233-G255, K326-V339, I432-R444, V457-L465</td><td>BLIMPS-PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel:
DM00560|P23574|26-465: L26-L465
DM00560|P20237|20-556: L26-V396, A437-L463
DM00560|P16305|4-443: D63-S377, A437-L463
DM00560|P08219|14-456: T65-L463</td><td>BLAST-DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Ion channel/postsynaptic membrane receptor
PD000153: N127-Y356, Q66-V286</td><td>BLAST-PRODOM</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22					Ion channel/postsynaptic membrane receptor PD000604: G403-L463	BLAST-PRODOM
23	7477105CD1	235	S151 S6 T218 T56 T57 T90		Transmembrane domains: S103-N123, I136-R162	HMMER
					Nucleoside transporter, equilibrative: PD006749: P63-L157 (P-value = 1.0e-07)	BLAST-PRODOM
24	7482079CD1	662	S10 S12 S137 S211 S323 S5 S564 T130 T19 T195 T281 T403 T499 T627 T657 T83 Y187	N17 N440 N517	Transmembrane domain: G412-Y430	HMMER
					K+ channel tetramerisation domain: S97-F203	HMMER-PFAM
					Ion transport protein: G263-L609	HMMER-PFAM
					Potassium channel signature PR00169: Q410-E433, F441-L463, G587-F613, E148-S167, P253-T281, H304-L327, F330-L350, L381-C407	BLIMPS-PRINTS
					Potassium channel CDRK: DM00436 | JH0595 | 144-307: K215-L390 DM00436 | P15387 | 136-299: R206-L381 DM00436 | P17970 | 386-549: I216-L390 DM00490 | P17970 | 268-384: A94-R200	BLAST-DOMO
					Voltage-gated potassium channel: PD000141: F330-S469, V570-K619, I572-I645	BLAST-PRODOM
25	55145506CD1	371	S113 S158 S194 S330 T151 T211 T323 T341 T63 T8	N127 N209	PROTEIN CHANNEL IONIC ION TRANSPORT VOLTAGEGATED P64 CHLORIDE INTRACELLULAR CHLORINE PD017366: A169-K355	BLAST_PRODOM
					CHLORINE CHANNEL PROTEIN P64 IONIC ION TRANSPORT VOLTAGEGATED TRANSMEMBRANE PHOSPHORYLATION PD118116: M1-Q125	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
26	5950519CD1	468	S105 S176 S23 S4 S56 T161 T170 T220 T308 T358 T466		Mitochondrial carrier proteins domain: M184-T276, H278-H369, G375-R468	HMMER_PFAM
					EF hand: R13-L41, R81-L109, Q117-H145	HMMER_PFAM
					Mitochondrial energy transfer proteins signature BL00215: V190-Q214, I425-G437	BLIMPS_BLOCKS
					Mitochondrial energy transfer proteins signature: K187-L241, V279-P331, I376-Q428	PROFILESCAN
					Mitochondrial carrier proteins signature PR00926: Q188-T201, T201-V215, G244-E264, T292-R310, Y335-L353, G383-Q405	BLIMPS_PRINTS
					Grave's disease carrier protein signature PR00928: P205-I225	BLIMPS_PRINTS
					PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117: Q273-L463, K187-A293	BLAST_PRODOM
					MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026 | S57544 | 26-107: V190-I270 DM00026 | P29518 | 233-310: V284-K360 DM00026 | S54495 | 534-620: F283-N361 DM00026 | Q01888 | 126-214: H278-N361	BLAST_DOMO
					EF hand motifs: D22-L34, D90-I102	MOTIFS
					Mitochondrial carrier proteins motif: P299-L307	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
27	1687189CB1	2229	2190-2229, 759-1660	70564238V1	1269	1824
				7251266F7 (PROSTMY01)	1	693
				70300023D1	2059	2229
				7749453F8 (NOSEDIN01)	410	1006
				70565215V1	1741	2207
				2416733F6 (HNT3AZT01)	1147	1688
				7711767J1 (TESTTUE02)	756	1205
28	7078207CB1	7610	1-5580	7070225H1 (BRAUTDR02)	6275	6895
				6911060J1 (PITUDIR01)	3289	3865
				6772031H1 (BRAUNOR01)	1	696
				71063183V1	2885	3400
				6253219H1 (LUNPTUT02)	1442	2064
				6893301H1 (BRAITDR03)	5879	6157
				71065860V1	2228	2898
				6763740H1 (BRAUNOR01)	3901	4568
				6765621H1 (BRAUNOR01)	3843	4506
				7467144H1 (LUNGNOE02)	3406	3914
				3767813H1 (BRSTNOT24)	5937	6250
				6953905H1 (BRAITDR02)	5169	5868
				6977243H1 (BRAHTDR04)	4508	5140
				8016696J1 (BMARTXE01)	2071	2870
				5964168H1 (BRATNOT05)	6042	6690
				6762808J1 (BRAUNOR01)	2849	3387
				6950389H1 (BRAITDR02)	725	1478
				4098906F8 (BRAITUT26)	6463	7085
				7757265J1 (SPLNTUE01)	660	1225
				5098681F8 (EPIMNON05)	7071	7610
				7179893H1 (BRAXDIC01)	6909	7418
				71969653V1	1561	2109
				6908865J1 (PITUDIR01)	4613	5241
				6893778J1 (BRAITDR03)	5243	5900
29	1560619CB1	2219	1-1659	6452362F8 (COLNDIC01)	1	553
				71597474V1	1539	2219
				71594784V1	1331	2107
				70683177V1	1272	1783
				70680523V1	738	1346
				71596281V1	364	920
30	2614283CB1	1280	415-559	60202200D1	385	843
				8097352H1 (EYERNOA01)	850	1280
				7432729H1 (PANCDIR02)	425	1006

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
30				7987760H1 (UTRSTUC01)	1	439
31	2667691CB1	2727	1-330	5753102H1 (LUNGNOT35)	1332	1994
				71100388V1	654	1359
				GBI.g8081479.smoosh	1	266
				7312933H1 (SINTNON02)	1968	2591
				70233893V1	266	752
				GBI.g9988362.smoosh	168	326
				8093950H1 (EYERNOA01)	741	1387
				7925641H2 (COLNTUS02)	2052	2727
				7346378H1 (SYNODIN02)	1395	2014
32	3211415CB1	1631	1-43, 1303-1631	70062244V1	704	1145
				5313185F8 (KIDETXS02)	1	719
				70059213V1	1170	1631
				70057909V1	1010	1547
33	4739923CB1	2673	1483-1785, 1-37	71982150V1	1187	1830
				71986856V1	1513	2125
				4567241F7 (HELATXT01)	2337	2673
				71983447V1	1401	1841
				7260030H1 (BRAWNOC01)	1889	2499
				7997955H1 (BRAITUC02)	143	745
				6265341H1 (MCLDTXN03)	1	212
				3767715T6 (BRSTNOT24)	613	1253
34	55030459CB1	3958	1-274, 837-1623, 3909-3958	55030219H1	1368	1996
				71992529V1	2796	3508
				71990982V1	2748	3435
				6343107T8 (LUNGDIS03)	1167	1585
				GNN.g7960408_000016_002.edit	1	198
				55030491J1	300	785
				55030459H1	809	1368
				71992886V1	2086	2783
				71989595V1	3248	3958
				71990326V1	1875	2578
				55109637H1	495	1288
				g4018506	48	528
35	6113039CB1	2000	856-1096	71721645V1	1296	2000
				6782480F9 (SINITMC01)	1	653
				71722719V1	411	1050
				71719834V1	1000	1641

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
36	7101781CB1	1997	1568-1778, 644-1026	FL7101781_g7939384_000014_g8131858	174	1543
				70925356V1	1321	1579
				3748173F6 (UTRSNOT18)	1	726
				70990502V1	1424	1997
37	7473036CB1	3069	1-1362, 2182-2313, 1436-1763	FL7473036_g9255974_000002_g2198815	1	2839
				5050192F6 (BRSTNOT33)	2484	3069
38	7476943CB1	2241	1-168, 1540-2241, 282-745	6392952F8 (PANCNON03)	1562	2241
				FL7476943_g7739804_000008_g3047402	660	1765
				55140014J1	1	879
39	8003355CB1	1593	1-38, 1173-1315	8003355H1 (MUSCTDC01)	28	620
				GNN.g7651721_000004_004	49	1593
				3292859H1 (BONRFET01)	1	248
40	3116448CB1	2121	358-692, 1-22	2378367F6 (ISLTNOT01)	1258	1771
				55136206J2	1	779
				5723184F6 (SEMVNOT05)	1487	2121
				70769061V1	1069	1669
				55136206H1	4	858
				7169977H1 (MCLRNOC01)	771	1092
41	622868CB1	1225	1-87	70501768V1	708	1225
				1851960F6 (LUNGFET03)	1	529
				70502134V1	624	1196
				70501182V1	470	1114
42	7476494CB1	2693	1-1295, 2361-2451	1382551F6 (BRAITUT08)	1	504
				FL7476494_g9438678_000004_g7688148_1_5-6	2035	2271
				7175426H1 (BRSTTMC01)	1537	2100
				GNN.g9438678_000004_002	1707	2642
				FL7476494_g9438678_000004_g7688148_1_6-7	2147	2435
				55116347J1	401	1094
				7757711H1 (SPLNTUE01)	762	1236
				FL7476494_g9438678_000004_g7688148_1_8-9	2436	2693
				7757711J1 (SPLNTUE01)	1204	1863
43	7477260CB1	3569	1-2249, 3310-3569	GNN.g8468993_000014_002.edit	3130	3569

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
43				5546177F8 (TESTNOC01)	2197	3093
				5313313F8 (KIDETXS02)	995	1522
				8011222H1 (NOSEDIC02)	1	767
				55089843H1 (PROTDNV21)	668	1007
				7227359H1 (BRAXTDR15)	2847	3383
				55120438J1	1497	2198
44	1963058CB1	3920	1-648, 2120-3920	6769623H1 (BRAUNOR01)	1745	2338
				7611869J1 (KIDCTME01)	3090	3825
				7696528H1 (KIDPTDE01)	2104	2636
				7001668H1 (HEALDIR01)	511	1014
				1963058R6 (BRSTNOT04)	3337	3920
				8174904H1 (FETANOA01)	964	1575
				6770724H1 (BRAUNOR01)	1587	2140
				3164103H1 (TLYMTXT04)	1329	1623
				7314310H1 (UTREDME02)	2753	3339
				2659167H1 (LUNGTUT09)	2421	2664
				7727654H1 (UTRCDIE01)	1	560
				7412125H1 (BONMTUE02)	631	1242
45	2395967CB1	1361	523-715	g4689801	944	1361
				2395967F6 (THP1AZT01)	642	1214
				71526782V1	1	612
				6411441H1 (UTREDIT10)	827	1359
				71469742V1	532	833
46	3586648CB1	1867	1-71, 1837-1867	2738605T6 (OVARNOT09)	1219	1825
				70855458V1	623	1267
				3586648F6 (293TF4T01)	1	575
				g1383637	1437	1867
				71224790V1	503	1044
47	7473396CB1	2211	1-2211	GNN.g9212516_1	1	2211
48	7476283CB1	1446	1053-1092, 1-265, 606-682, 1140-1192	GBI.g7684447_12_11_07_09_05_10.edit	49	1446
				55110089J1	1	257
				55110065J1	745	1184
49	7477105CB1	1332	1-819	71223112V1	434	1069
				7948245J1 (BRABNOE02)	1	512
				71040789V1	549	1156
				6711669H1 (BRABDIT01)	968	1332
50	7482079CB1	2298	1-732, 1302-1712, 861-895	GNN.g9650542_2	1	1989

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
50				g3765560	1929	2298
51	55145506CB1	2250	1-555, 1490-1754, 1320-1374, 1094-1143	72396051V1	1286	1935
				72393047V1	1615	2250
				70771274V1	1235	1822
				55145606J1	1	660
				70772827V1	717	1293
				72394339V1	610	1245
52	5950519CB1	3430	1-35, 3250-3430, 3109-3130, 2255-2277	2106229T6 (BRAITUT03)	2926	3404
				70378849D1	1595	2198
				7096023H1 (BRACDIR02)	2177	2851
				6327536H1 (BRANDIN01)	2987	3430
				6764621H1 (BRAUNOR01)	1331	1899
				6764621J1 (BRAUNOR01)	1	708
				6307874H1 (NERDTDN03)	668	1258
				6980581H1 (BRAHTDR04)	1177	1527
				6121921H1 (BRAHNON05)	2426	2986

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
27	1687189CB1	PROSTMY01
28	7078207CB1	BRAUNOR01
29	1560619CB1	LUNGNOT37
30	2614283CB1	PROSTUT09
31	2667691CB1	STOMFET01
32	3211415CB1	BLADNOT08
33	4739923CB1	BRAITUT03
34	55030459CB1	BRAYDIN03
35	6113039CB1	SINITMC01
36	7101781CB1	LUNGNOT34
37	7473036CB1	BRSTNOT33
38	7476943CB1	PANCNON03
39	8003355CB1	BONRFET01
40	3116448CB1	SEMVNOT05
41	622868CB1	PGANNOT01
42	7476494CB1	SPLNTUE01
43	7477260CB1	TESTNOC01
44	1963058CB1	BRAUNOR01
45	2395967CB1	THP1AZT01
46	3586648CB1	OVARNOT09
49	7477105CB1	COLNNOT11
51	55145506CB1	SINITMR01
52	5950519CB1	BRAUNOR01

Table 6

Library	Vector	Library Description
BLADNOT08	pINCY	Library was constructed using RNA isolated from the bladder tissue of an 11-year-old black male, who died from a gunshot wound.
BONRFET01	pINCY	Library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.
BRAITUT03	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select

Table 6 (cont.)

Library	Vector	Library Description
		for insert containing clones.
BRSTNOT33	pINCY	Library was constructed using RNA isolated from right breast tissue removed from a 46-year-old Caucasian female during unilateral extended simple mastectomy with breast reconstruction. Pathology for the associated tumor tissue indicated invasive grade 3 adenocarcinoma, ductal type, with apocrine features, nuclear grade 3 forming a mass in the outer quadrant. There was greater than 50% intraductal component. Patient history included breast cancer.
COLNNOT11	PSPORT1	Library was constructed using RNA isolated from colon tissue removed from a 60-year-old Caucasian male during a left hemicolectomy.
LUNGNOT34	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 12-year-old Caucasian male.
LUNGNOT37	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 15-year-old Caucasian female who died from a closed head injury. Serology was positive for cytomegalovirus.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
PANCNON03	pINCY	This normalized pancreas tissue library was constructed from 12 million independent clones from a pancreas library. Starting RNA was made from RNA isolated from pancreas tissue removed from a 17-year-old Caucasian female who died from head trauma. Serology was positive for cytomegalovirus and remaining serologies were negative. The patient was not taking any medications. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PGANNOT01	PSPORT1	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.
PROSTMY01	pINCY	This large size-fractionated cDNA and normalized library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node

Table 6 (cont.)

Library	Vector	Library Description
		excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma Gleason grade 4 forming a predominant mass involving the left side peripherally with extension into the right posterior superior region. The tumor invaded the capsule and perforated the capsule to involve periprostatic tissue in the left posterior superior region. The left inferior posterior and left superior posterior surgical margins are positive. One left pelvic lymph node is metastatically involved. Patient history included calculus of the kidney. Family history included lung cancer and breast cancer. The size-selected library was normalized in 1 round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
SEMVNOT05	pINCY	Library was constructed using RNA isolated from seminal vesicle tissue removed from a 67-year-old Caucasian male during radical prostatectomy. Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason grade 3 + 3.
SINITMC01	pINCY	This large size-fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from ileum tissue removed from a 30-year-old Caucasian female (donor A) during partial colectomy, open liver biopsy, and permanent colostomy, and from ileum tissue removed from a 70-year-old Caucasian female (donor B) during right hemicolectomy, open liver biopsy, sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue (donor A) indicated carcinoid tumor (grade 1 neuroendocrine carcinoma) arising in the terminal ileum. The tumor permeated through the ileal wall into the mesenteric fat and extended into the adherent cecum, where tumor extended through the bowel wall up to the mucosal surface. Multiple lymph nodes were positive for tumor. Additional (2) lymph nodes were also involved by direct tumor extension. Pathology for donor B indicated a non-tumorous margin of ileum. Pathology for the matched tumor (donor B) indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated distal to the ileocecal valve. The tumor invaded through the muscularis propria just into the serosal adipose tissue. One regional lymph node was positive for a microfocus of metastatic

Table 6 (cont.)

Library	Vector	Library Description
		adenocarcinoma. Donor A presented with flushing and unspecified abdominal/pelvic symptoms. Patient history included endometriosis, and tobacco and alcohol abuse. Donor B's history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, and a malignant skin neoplasm. Donor B's medication included tamoxifen.
SINITMR01	PCDNA2.1	This random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.
SPLNTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from spleen tumor tissue removed from a 28-year-old male during total splenectomy. Pathology indicated malignant lymphoma, diffuse large cell type, B-cell phenotype with abundant reactive T-cells and marked granulomatous response involving the spleen, where it formed approximately 45 nodules, liver, and multiple lymph nodes.
STOMFET01	pINCY	Library was constructed using RNA isolated from the stomach tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
THP1AZT01	pINCY	Library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-26.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:27-52.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

19. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expressi n library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in the sample.

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide

or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence f SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

<110> INCYTE GENOMICS, INC.
LEE, Ernestine A.
YUE, Henry
LAL, Preeti G.
WALIA, Narinder K.
BAUGHN, Mariah R.
WARREN, Bridget A.
LEE, Sally
SANJANWALA, Madhu S.
YAO, Monique G.
RAMKUMAR, Jayalaxmi
THORNTON, Michael
GANDHI, Ameena R.
POLICKY, Jennifer L.
ELLIOTT, Vicki S.
ARVIZU, Chandra
RAUMANN, Brigitte E.
BRUNS, Christopher M.
NAINA, Amir
HAFALIA, April J.A.
NGUYEN, Danniel B.
XU, Yuming
LU, Dyung Aina M.
ISON, Craig H.
GRIFFIN, Jennifer A.
REDDY, Roopa M.
BURFORD, Neil

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0217 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/232,685; 60/234,842; 60/236,882; 60/239,057; 60/240,540; 60/241,700
<151> 2000-09-15; 2000-09-22; 2000-09-29; 2000-10-05; 2000-13-10; 2000-10-18

<160> 52
<170> PERL Program

<210> 1
<211> 619
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1687189CD1

<400> 1
Met Pro Trp Gln Ala Phe Arg Arg Phe Gly Gln Lys Leu Val Arg
1 5 10 15
Arg Arg Thr Leu Glu Ser Gly Met Ala Glu Thr Arg Leu Ala Arg

```
                20                  25                  30
Cys Leu Ser Thr Leu Asp Leu Val Ala Leu Gly Val Gly Ser Thr
                35                  40                  45
Leu Gly Ala Gly Val Tyr Val Leu Ala Gly Glu Val Ala Lys Asp
                50                  55                  60
Lys Ala Gly Pro Ser Ile Val Ile Cys Phe Leu Val Ala Ala Leu
                65                  70                  75
Ser Ser Val Leu Ala Gly Leu Cys Tyr Ala Glu Phe Gly Ala Arg
                80                  85                  90
Val Pro Arg Ser Gly Ser Ala Tyr Leu Tyr Ser Tyr Val Thr Val
                95                  100                 105
Gly Glu Leu Trp Ala Phe Thr Thr Gly Trp Asn Leu Ile Leu Ser
                110                 115                 120
Tyr Val Ile Gly Thr Ala Ser Val Ala Arg Ala Trp Ser Ser Ala
                125                 130                 135
Phe Asp Asn Leu Ile Gly Asn His Ile Ser Lys Thr Leu Gln Gly
                140                 145                 150
Ser Ile Ala Leu His Val Pro His Val Leu Ala Glu Tyr Pro Asp
                155                 160                 165
Phe Phe Ala Leu Gly Leu Val Leu Leu Leu Thr Gly Leu Leu Ala
                170                 175                 180
Leu Gly Ala Ser Glu Ser Ala Leu Val Thr Lys Val Phe Thr Gly
                185                 190                 195
Val Asn Leu Leu Val Leu Gly Phe Val Met Ile Ser Gly Phe Val
                200                 205                 210
Lys Gly Asp Val His Asn Trp Lys Leu Thr Glu Glu Asp Tyr Glu
                215                 220                 225
Leu Ala Met Ala Glu Leu Asn Asp Thr Tyr Ser Leu Gly Pro Leu
                230                 235                 240
Gly Ser Gly Gly Phe Val Pro Phe Gly Phe Glu Gly Ile Leu Arg
                245                 250                 255
Gly Ala Ala Thr Cys Phe Tyr Ala Phe Val Gly Phe Asp Cys Ile
                260                 265                 270
Ala Thr Thr Gly Glu Glu Ala Gln Asn Pro Gln Arg Ser Ile Pro
                275                 280                 285
Met Gly Ile Val Ile Ser Leu Ser Val Cys Phe Leu Ala Tyr Phe
                290                 295                 300
Ala Val Ser Ser Ala Leu Thr Leu Met Met Pro Tyr Tyr Gln Leu
                305                 310                 315
Gln Pro Glu Ser Pro Leu Pro Glu Ala Phe Leu Tyr Ile Gly Trp
                320                 325                 330
Ala Pro Ala Arg Tyr Val Val Ala Val Gly Ser Leu Cys Ala Leu
                335                 340                 345
Ser Thr Ser Leu Leu Gly Ser Met Phe Pro Met Pro Arg Val Ile
                350                 355                 360
Tyr Ala Met Ala Glu Asp Gly Leu Leu Phe Arg Val Leu Ala Arg
                365                 370                 375
Ile His Thr Gly Thr Arg Thr Pro Ile Ile Ala Thr Val Val Ser
                380                 385                 390
Gly Ile Ile Ala Ala Phe Met Ala Phe Leu Phe Lys Leu Thr Asp
                395                 400                 405
Leu Val Asp Leu Met Ser Ile Gly Thr Leu Leu Ala Tyr Ser Leu
                410                 415                 420
Val Ser Ile Cys Val Leu Ile Leu Arg Tyr Gln Pro Asp Gln Glu
                425                 430                 435
Thr Lys Thr Gly Glu Glu Val Glu Leu Gln Glu Glu Ala Ile Thr
```

```
                440                 445                 450
Thr Glu Ser Glu Lys Leu Thr Leu Trp Gly Leu Phe Phe Pro Leu
                455                 460                 465
Asn Ser Ile Pro Thr Pro Leu Ser Gly Gln Ile Val Tyr Val Cys
                470                 475                 480
Ser Ser Leu Leu Ala Val Leu Leu Thr Ala Leu Cys Leu Val Leu
                485                 490                 495
Ala Gln Trp Ser Val Pro Leu Leu Ser Gly Asp Leu Leu Trp Thr
                500                 505                 510
Ala Val Val Val Leu Leu Leu Leu Leu Ile Ile Gly Ile Ile Val
                515                 520                 525
Val Ile Trp Arg Gln Pro Gln Ser Ser Thr Pro Leu His Phe Lys
                530                 535                 540
Val Pro Ala Leu Pro Leu Leu Pro Leu Met Ser Ile Phe Val Asn
                545                 550                 555
Ile Tyr Leu Met Met Gln Met Thr Ala Gly Thr Trp Ala Arg Phe
                560                 565                 570
Gly Val Trp Met Leu Ile Gly Phe Ala Ile Tyr Phe Gly Tyr Gly
                575                 580                 585
Ile Gln His Ser Leu Glu Glu Ile Lys Ser Asn Gln Pro Ser Arg
                590                 595                 600
Lys Ser Arg Ala Lys Thr Val Asp Leu Asp Pro Gly Thr Leu Tyr
                605                 610                 615
Val His Ser Val


<210> 2
<211> 2436
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7078207CD1

<400> 2
Met Gly Phe Leu His Gln Leu Gln Leu Leu Leu Trp Lys Asn Val
  1             5                  10                  15
Thr Leu Lys Arg Arg Ser Pro Trp Val Leu Ala Phe Glu Ile Phe
                20                  25                  30
Ile Pro Leu Val Leu Phe Phe Ile Leu Leu Gly Leu Arg Gln Lys
                35                  40                  45
Lys Pro Thr Ile Ser Val Lys Glu Val Ser Phe Tyr Thr Ala Ala
                50                  55                  60
Pro Leu Thr Ser Ala Gly Ile Leu Pro Val Met Gln Ser Leu Cys
                65                  70                  75
Pro Asp Gly Gln Arg Asp Glu Phe Gly Phe Leu Gln Tyr Ala Asn
                80                  85                  90
Ser Thr Val Thr Gln Leu Leu Glu Arg Leu Asp Arg Val Val Glu
                95                 100                 105
Glu Gly Asn Leu Phe Asp Pro Ala Arg Pro Ser Leu Gly Ser Glu
               110                 115                 120
Leu Glu Ala Leu Arg Gln His Leu Glu Ala Leu Ser Ala Gly Pro
               125                 130                 135
Gly Thr Ser Gly Ser His Leu Asp Arg Ser Thr Val Ser Ser Phe
               140                 145                 150
```

Ser Leu Asp Ser Val Ala Arg Asn Pro Gln Glu Leu Trp Arg Phe
155 160 165
Leu Thr Gln Asn Leu Ser Leu Pro Asn Ser Thr Ala Gln Ala Leu
170 175 180
Leu Ala Ala Arg Val Asp Pro Pro Glu Val Tyr His Leu Leu Phe
185 190 195
Gly Pro Ser Ser Ala Leu Asp Ser Gln Ser Gly Leu His Lys Gly
200 205 210
Gln Glu Pro Trp Ser Arg Leu Gly Gly Asn Pro Leu Phe Arg Met
215 220 225
Glu Glu Leu Leu Leu Ala Pro Ala Leu Leu Glu Gln Leu Thr Cys
230 235 240
Thr Pro Gly Ser Gly Glu Leu Gly Arg Ile Leu Thr Val Pro Glu
245 250 255
Ser Gln Lys Gly Ala Leu Gln Gly Tyr Arg Asp Ala Val Cys Ser
260 265 270
Gly Gln Ala Ala Ala Arg Ala Arg Arg Phe Ser Gly Leu Ser Ala
275 280 285
Glu Leu Arg Asn Gln Leu Asp Val Ala Lys Val Ser Gln Gln Leu
290 295 300
Gly Leu Asp Ala Pro Asn Gly Ser Asp Ser Ser Pro Gln Ala Pro
305 310 315
Pro Pro Arg Arg Leu Gln Ala Leu Leu Gly Asp Leu Leu Asp Ala
320 325 330
Gln Lys Val Leu Gln Asp Val Asp Val Leu Ser Ala Leu Ala Leu
335 340 345
Leu Leu Pro Gln Gly Ala Cys Thr Gly Arg Thr Pro Gly Pro Pro
350 355 360
Ala Ser Gly Ala Gly Gly Ala Ala Asn Gly Thr Gly Ala Gly Ala
365 370 375
Val Met Gly Pro Asn Ala Thr Ala Glu Glu Gly Ala Pro Ser Ala
380 385 390
Ala Ala Leu Ala Thr Pro Asp Thr Leu Gln Gly Gln Cys Ser Ala
395 400 405
Phe Val Gln Leu Trp Ala Gly Leu Gln Pro Ile Leu Cys Gly Asn
410 415 420
Asn Arg Thr Ile Glu Pro Glu Ala Leu Arg Arg Gly Asn Met Ser
425 430 435
Ser Leu Gly Phe Thr Ser Lys Glu Gln Arg Asn Leu Gly Leu Leu
440 445 450
Val His Leu Met Thr Ser Asn Pro Lys Ile Leu Tyr Ala Pro Ala
455 460 465
Gly Ser Glu Val Asp Arg Val Ile Leu Lys Ala Asn Glu Thr Phe
470 475 480
Ala Phe Val Gly Asn Val Thr His Tyr Ala Gln Val Trp Leu Asn
485 490 495
Ile Ser Ala Glu Ile Arg Ser Phe Leu Glu Gln Gly Arg Leu Gln
500 505 510
Gln His Leu Arg Trp Leu Gln Gln Tyr Val Ala Glu Leu Arg Leu
515 520 525
His Pro Glu Ala Leu Asn Leu Ser Leu Asp Glu Leu Pro Pro Ala
530 535 540
Leu Arg Gln Asp Asn Phe Ser Leu Pro Ser Gly Met Ala Leu Leu
545 550 555
Gln Gln Leu Asp Thr Ile Asp Asn Ala Ala Cys Gly Trp Ile Gln
560 565 570

Phe Met Ser Lys Val Ser Val Asp Ile Phe Lys Gly Phe Pro Asp
575 580 585
Glu Glu Ser Ile Val Asn Tyr Thr Leu Asn Gln Ala Tyr Gln Asp
590 595 600
Asn Val Thr Val Phe Ala Ser Val Ile Phe Gln Thr Arg Lys Asp
605 610 615
Gly Ser Leu Pro Pro His Val His Tyr Lys Ile Arg Gln Asn Ser
620 625 630
Ser Phe Thr Glu Lys Thr Asn Glu Ile Arg Arg Ala Tyr Trp Arg
635 640 645
Pro Gly Pro Asn Thr Gly Gly Arg Phe Tyr Phe Leu Tyr Gly Phe
650 655 660
Val Trp Ile Gln Asp Met Met Glu Arg Ala Ile Ile Asp Thr Phe
665 670 675
Val Gly His Asp Val Val Glu Pro Gly Ser Tyr Val Gln Met Phe
680 685 690
Pro Tyr Pro Cys Tyr Thr Arg Asp Asp Phe Leu Phe Val Ile Glu
695 700 705
His Met Met Pro Leu Cys Met Val Ile Ser Trp Val Tyr Ser Val
710 715 720
Ala Met Thr Ile Gln His Ile Val Ala Glu Lys Glu His Arg Leu
725 730 735
Lys Glu Val Met Lys Thr Met Gly Leu Asn Asn Ala Val His Trp
740 745 750
Val Ala Trp Phe Ile Thr Gly Phe Val Gln Leu Ser Ile Ser Val
755 760 765
Thr Ala Leu Thr Ala Ile Leu Lys Tyr Gly Gln Val Leu Met His
770 775 780
Ser His Val Val Ile Ile Trp Leu Phe Leu Ala Val Tyr Ala Val
785 790 795
Ala Thr Ile Met Phe Cys Phe Leu Val Ser Val Leu Tyr Ser Lys
800 805 810
Ala Lys Leu Ala Ser Ala Cys Gly Gly Ile Ile Tyr Phe Leu Ser
815 820 825
Tyr Val Pro Tyr Met Tyr Val Ala Ile Arg Glu Glu Val Ala His
830 835 840
Asp Lys Ile Thr Ala Phe Glu Lys Cys Ile Ala Ser Leu Met Ser
845 850 855
Thr Thr Ala Phe Gly Leu Gly Ser Lys Tyr Phe Ala Leu Tyr Glu
860 865 870
Val Ala Gly Val Gly Ile Gln Trp His Thr Phe Ser Gln Ser Pro
875 880 885
Val Glu Gly Asp Asp Phe Asn Leu Leu Leu Ala Val Thr Met Leu
890 895 900
Met Val Asp Ala Val Val Tyr Gly Ile Leu Thr Trp Tyr Ile Glu
905 910 915
Ala Val His Pro Gly Met Tyr Gly Leu Pro Arg Pro Trp Tyr Phe
920 925 930
Pro Leu Gln Lys Ser Tyr Trp Leu Gly Ser Gly Arg Thr Glu Ala
935 940 945
Trp Glu Trp Ser Trp Pro Trp Ala Arg Thr Pro Arg Leu Ser Val
950 955 960
Met Glu Glu Asp Gln Ala Cys Ala Met Glu Ser Arg Arg Phe Glu
965 970 975
Glu Thr Arg Gly Met Glu Glu Glu Pro Thr His Leu Pro Leu Val
980 985 990

Val Cys Val Asp Lys Leu Thr Lys Val Tyr Lys Asp Asp Lys Lys
995 1000 1005
Leu Ala Leu Asn Lys Leu Ser Leu Asn Leu Tyr Glu Asn Gln Val
1010 1015 1020
Val Ser Phe Leu Gly His Asn Gly Ala Gly Lys Thr Thr Thr Met
1025 1030 1035
Ser Ile Leu Thr Gly Leu Phe Pro Pro Thr Ser Gly Ser Ala Thr
1040 1045 1050
Ile Tyr Gly His Asp Ile Arg Thr Glu Met Asp Glu Ile Arg Lys
1055 1060 1065
Asn Leu Gly Met Cys Pro Gln His Asn Val Leu Phe Asp Arg Leu
1070 1075 1080
Thr Val Glu Glu His Leu Trp Phe Tyr Ser Arg Leu Lys Ser Met
1085 1090 1095
Ala Gln Glu Glu Ile Arg Arg Glu Met Asp Lys Met Ile Glu Asp
1100 1105 1110
Leu Glu Leu Ser Asn Lys Arg His Ser Leu Val Gln Thr Leu Ser
1115 1120 1125
Gly Gly Met Lys Arg Lys Leu Ser Val Ala Ile Ala Phe Val Gly
1130 1135 1140
Gly Ser Arg Ala Ile Ile Leu Asp Glu Pro Thr Ala Gly Val Asp
1145 1150 1155
Pro Tyr Ala Arg Arg Ala Ile Trp Asp Leu Ile Leu Lys Tyr Lys
1160 1165 1170
Pro Gly Arg Thr Ile Leu Leu Ser Thr His His Met Asp Glu Ala
1175 1180 1185
Asp Leu Leu Gly Asp Arg Ile Ala Ile Ile Ser His Gly Lys Leu
1190 1195 1200
Lys Cys Cys Gly Ser Pro Leu Phe Leu Lys Gly Thr Tyr Gly Asp
1205 1210 1215
Gly Tyr Arg Leu Thr Leu Val Lys Arg Pro Ala Glu Pro Gly Gly
1220 1225 1230
Pro Gln Glu Pro Gly Leu Ala Ser Ser Pro Pro Gly Arg Ala Pro
1235 1240 1245
Leu Ser Ser Cys Ser Glu Leu Gln Val Ser Gln Phe Ile Arg Lys
1250 1255 1260
His Val Ala Ser Cys Leu Leu Val Ser Asp Thr Ser Thr Glu Leu
1265 1270 1275
Ser Tyr Ile Leu Pro Ser Glu Ala Ala Lys Lys Gly Ala Phe Glu
1280 1285 1290
Arg Leu Phe Gln His Leu Glu Arg Ser Leu Asp Ala Leu His Leu
1295 1300 1305
Ser Ser Phe Gly Leu Met Asp Thr Thr Leu Glu Glu Val Phe Leu
1310 1315 1320
Lys Val Ser Glu Glu Asp Gln Ser Leu Glu Asn Ser Glu Ala Asp
1325 1330 1335
Val Lys Glu Ser Arg Lys Asp Val Leu Pro Gly Ala Glu Gly Pro
1340 1345 1350
Ala Ser Gly Glu Gly His Ala Gly Asn Leu Ala Arg Cys Ser Glu
1355 1360 1365
Leu Thr Gln Ser Gln Ala Ser Leu Gln Ser Ala Ser Ser Val Gly
1370 1375 1380
Ser Ala Arg Gly Asp Glu Gly Ala Gly Tyr Thr Asp Val Tyr Gly
1385 1390 1395
Asp Tyr Arg Pro Leu Phe Asp Asn Pro Gln Asp Pro Asp Asn Val
1400 1405 1410

Ser Leu Gln Glu Val Glu Ala Glu Ala Leu Ser Arg Val Gly Gln
1415 1420 1425
Gly Ser Arg Lys Leu Asp Gly Gly Trp Leu Lys Val Arg Gln Phe
1430 1435 1440
His Gly Leu Leu Val Lys Arg Phe His Cys Ala Arg Arg Asn Ser
1445 1450 1455
Lys Ala Leu Phe Ser Gln Ile Leu Leu Pro Ala Phe Phe Val Cys
1460 1465 1470
Val Ala Met Thr Val Ala Leu Ser Val Pro Glu Ile Gly Asp Leu
1475 1480 1485
Pro Pro Leu Val Leu Ser Pro Ser Gln Tyr His Asn Tyr Thr Gln
1490 1495 1500
Pro Arg Gly Asn Phe Ile Pro Tyr Ala Asn Glu Glu Arg Arg Glu
1505 1510 1515
Tyr Arg Leu Arg Leu Ser Pro Asp Ala Ser Pro Gln Gln Leu Val
1520 1525 1530
Ser Thr Phe Arg Leu Pro Ser Gly Val Gly Ala Thr Cys Val Leu
1535 1540 1545
Lys Ser Pro Ala Asn Gly Ser Leu Gly Pro Thr Leu Asn Leu Ser
1550 1555 1560
Ser Gly Glu Ser Arg Leu Leu Ala Ala Arg Phe Phe Asp Ser Met
1565 1570 1575
Cys Leu Glu Ser Phe Thr Gln Gly Leu Pro Leu Ser Asn Phe Val
1580 1585 1590
Pro Pro Pro Pro Ser Pro Ala Pro Ser Asp Ser Pro Ala Ser Pro
1595 1600 1605
Asp Glu Asp Leu Gln Ala Trp Asn Val Ser Leu Pro Pro Thr Ala
1610 1615 1620
Gly Pro Glu Met Trp Thr Ser Ala Pro Ser Leu Pro Arg Leu Val
1625 1630 1635
Arg Glu Pro Val Arg Cys Thr Cys Ser Ala Gln Gly Thr Gly Phe
1640 1645 1650
Ser Cys Pro Ser Ser Val Gly Gly His Pro Pro Gln Met Arg Val
1655 1660 1665
Val Thr Gly Asp Ile Leu Thr Asp Ile Thr Gly His Asn Val Ser
1670 1675 1680
Glu Tyr Leu Leu Phe Thr Ser Asp Arg Phe Arg Leu His Arg Tyr
1685 1690 1695
Gly Ala Ile Thr Phe Gly Asn Val Leu Lys Ser Ile Pro Ala Ser
1700 1705 1710
Phe Gly Thr Arg Ala Pro Pro Met Val Arg Lys Ile Ala Val Arg
1715 1720 1725
Arg Ala Ala Gln Val Phe Tyr Asn Asn Lys Gly Tyr His Ser Met
1730 1735 1740
Pro Thr Tyr Leu Asn Ser Leu Asn Asn Ala Ile Leu Arg Ala Asn
1745 1750 1755
Leu Pro Lys Ser Lys Gly Asn Pro Ala Ala Tyr Gly Ile Thr Val
1760 1765 1770
Thr Asn His Pro Met Asn Lys Thr Ser Ala Ser Leu Ser Leu Asp
1775 1780 1785
Tyr Leu Leu Gln Gly Thr Asp Val Val Ile Ala Ile Phe Ile Ile
1790 1795 1800
Val Ala Met Ser Phe Val Pro Ala Ser Phe Val Val Phe Leu Val
1805 1810 1815
Ala Glu Lys Ser Thr Lys Ala Lys His Leu Gln Phe Val Ser Gly
1820 1825 1830

Cys Asn Pro Ile Ile Tyr Trp Leu Ala Asn Tyr Val Trp Asp Met
1835 1840 1845
Leu Asn Tyr Leu Val Pro Ala Thr Cys Cys Val Ile Ile Leu Phe
1850 1855 1860
Val Phe Asp Leu Pro Ala Tyr Thr Ser Pro Thr Asn Phe Pro Ala
1865 1870 1875
Val Leu Ser Leu Phe Leu Leu Tyr Gly Trp Ser Ile Thr Pro Ile
1880 1885 1890
Met Tyr Pro Ala Ser Phe Trp Phe Glu Val Pro Ser Ser Ala Tyr
1895 1900 1905
Val Phe Leu Ile Val Ile Asn Leu Phe Ile Gly Ile Thr Ala Thr
1910 1915 1920
Val Ala Thr Phe Leu Leu Gln Leu Phe Glu His Asp Lys Asp Leu
1925 1930 1935
Lys Val Val Asn Ser Tyr Leu Lys Ser Cys Phe Leu Ile Phe Pro
1940 1945 1950
Asn Tyr Asn Leu Gly His Gly Leu Met Glu Met Ala Tyr Asn Glu
1955 1960 1965
Tyr Ile Asn Glu Tyr Tyr Ala Lys Ile Gly Gln Phe Asp Lys Met
1970 1975 1980
Lys Ser Pro Phe Glu Trp Asp Ile Val Thr Arg Gly Leu Val Ala
1985 1990 1995
Met Ala Val Glu Gly Val Val Gly Phe Leu Leu Thr Ile Met Cys
2000 2005 2010
Gln Tyr Asn Phe Leu Arg Arg Pro Gln Arg Met Pro Val Ser Thr
2015 2020 2025
Lys Pro Val Glu Asp Asp Val Asp Val Ala Ser Glu Arg Gln Arg
2030 2035 2040
Val Leu Arg Gly Asp Ala Asp Asn Asp Met Val Lys Ile Glu Asn
2045 2050 2055
Leu Thr Lys Val Tyr Lys Ser Arg Lys Ile Gly Arg Ile Leu Ala
2060 2065 2070
Val Asp Arg Leu Cys Leu Gly Val Arg Pro Gly Glu Cys Phe Gly
2075 2080 2085
Leu Leu Gly Val Asn Gly Ala Gly Lys Thr Ser Thr Phe Lys Met
2090 2095 2100
Leu Thr Gly Asp Glu Ser Thr Thr Gly Gly Glu Ala Phe Val Asn
2105 2110 2115
Gly His Ser Val Leu Lys Glu Leu Leu Gln Val Gln Gln Ser Leu
2120 2125 2130
Gly Tyr Cys Pro Gln Cys Asp Ala Leu Phe Asp Glu Leu Thr Ala
2135 2140 2145
Arg Glu His Leu Gln Leu Tyr Thr Arg Leu Arg Gly Ile Ser Trp
2150 2155 2160
Lys Asp Glu Ala Arg Val Val Lys Trp Ala Leu Glu Lys Leu Glu
2165 2170 2175
Leu Thr Lys Tyr Ala Asp Lys Pro Ala Gly Thr Tyr Ser Gly Gly
2180 2185 2190
Asn Lys Arg Lys Leu Ser Thr Ala Ile Ala Leu Ile Gly Tyr Pro
2195 2200 2205
Ala Phe Ile Phe Leu Asp Glu Pro Thr Thr Gly Met Asp Pro Lys
2210 2215 2220
Ala Arg Arg Phe Leu Trp Asn Leu Ile Leu Asp Leu Ile Lys Thr
2225 2230 2235
Gly Arg Ser Val Val Leu Thr Ser His Ser Met Glu Glu Cys Glu
2240 2245 2250

Ala Leu Cys Thr Arg Leu Ala Ile Met Val Asn Gly Arg Leu Arg
2255 2260 2265
Cys Leu Gly Ser Ile Gln His Leu Lys Asn Arg Phe Gly Asp Gly
2270 2275 2280
Tyr Met Ile Thr Val Arg Thr Lys Ser Ser Gln Ser Val Lys Asp
2285 2290 2295
Val Val Arg Phe Phe Asn Arg Asn Phe Pro Glu Ala Met Leu Lys
2300 2305 2310
Glu Arg His His Thr Lys Val Gln Tyr Gln Leu Lys Ser Glu His
2315 2320 2325
Ile Ser Leu Ala Gln Val Phe Ser Lys Met Glu Gln Val Ser Gly
2330 2335 2340
Val Leu Gly Ile Glu Asp Tyr Ser Val Ser Gln Thr Thr Leu Asp
2345 2350 2355
Asn Val Phe Val Asn Phe Ala Lys Lys Gln Ser Asp Asn Leu Glu
2360 2365 2370
Gln Gln Glu Thr Glu Pro Pro Ser Ala Leu Gln Ser Pro Leu Gly
2375 2380 2385
Cys Leu Leu Ser Leu Leu Arg Pro Arg Ser Ala Pro Thr Glu Leu
2390 2395 2400
Arg Ala Leu Val Ala Asp Glu Pro Glu Asp Leu Asp Thr Glu Asp
2405 2410 2415
Glu Gly Leu Ile Ser Phe Glu Glu Glu Arg Ala Gln Leu Ser Phe
2420 2425 2430
Asn Thr Asp Thr Leu Cys
2435

<210> 3
<211> 610
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1560619CD1

<400> 3
Met Ser Arg Ser Pro Leu Asn Pro Ser Gln Leu Arg Ser Val Gly
1 5 10 15
Ser Gln Asp Ala Leu Ala Pro Leu Pro Pro Pro Ala Pro Gln Asn
20 25 30
Pro Ser Thr His Ser Trp Asp Pro Leu Cys Gly Ser Leu Pro Trp
35 40 45
Gly Leu Ser Cys Leu Leu Ala Leu Gln His Val Leu Val Met Ala
50 55 60
Ser Leu Leu Cys Val Ser His Leu Leu Leu Leu Cys Ser Leu Ser
65 70 75
Pro Gly Gly Leu Ser Tyr Ser Pro Ser Gln Leu Leu Ala Ser Ser
80 85 90
Phe Phe Ser Cys Gly Met Ser Thr Ile Leu Gln Thr Trp Met Gly
95 100 105
Ser Arg Leu Pro Leu Val Gln Ala Pro Ser Leu Glu Phe Leu Ile
110 115 120
Pro Ala Leu Val Leu Thr Ser Gln Lys Leu Pro Arg Ala Ile Gln
125 130 135
Thr Pro Gly Asn Ser Ser Leu Met Leu His Leu Cys Arg Gly Pro

```
                140                 145                 150
Ser Cys His Gly Leu Gly His Trp Asn Thr Ser Leu Gln Glu Val
                155                 160                 165
Ser Gly Ala Val Val Val Ser Gly Leu Leu Gln Gly Met Met Gly
                170                 175                 180
Leu Leu Gly Ser Pro Gly His Val Phe Pro His Cys Gly Pro Leu
                185                 190                 195
Val Leu Ala Pro Ser Leu Val Val Ala Gly Leu Ser Ala His Arg
                200                 205                 210
Glu Val Ala Gln Phe Cys Phe Thr His Trp Gly Leu Ala Leu Leu
                215                 220                 225
Val Ile Leu Leu Met Val Val Cys Ser Gln His Leu Gly Ser Cys
                230                 235                 240
Gln Phe His Val Cys Pro Trp Arg Arg Ala Ser Thr Ser Ser Thr
                245                 250                 255
His Thr Pro Leu Pro Val Phe Arg Leu Leu Ser Val Leu Ile Pro
                260                 265                 270
Val Ala Cys Val Trp Ile Val Ser Ala Phe Val Gly Phe Ser Val
                275                 280                 285
Ile Pro Gln Glu Leu Ser Ala Pro Thr Lys Ala Pro Trp Ile Trp
                290                 295                 300
Leu Pro His Pro Gly Glu Trp Asn Trp Pro Leu Leu Thr Pro Arg
                305                 310                 315
Ala Leu Ala Ala Gly Ile Ser Met Ala Leu Ala Ala Ser Thr Ser
                320                 325                 330
Ser Leu Gly Cys Tyr Ala Leu Cys Gly Arg Leu Leu His Leu Pro
                335                 340                 345
Pro Pro Pro Pro His Ala Cys Ser Arg Gly Leu Ser Leu Glu Gly
                350                 355                 360
Leu Gly Ser Val Leu Ala Gly Leu Leu Gly Ser Pro Met Gly Thr
                365                 370                 375
Ala Ser Ser Phe Pro Asn Val Gly Lys Val Gly Leu Ile Gln Ala
                380                 385                 390
Gly Ser Gln Gln Val Ala His Leu Val Gly Leu Leu Cys Val Gly
                395                 400                 405
Leu Gly Leu Ser Pro Arg Leu Ala Gln Leu Leu Thr Thr Ile Pro
                410                 415                 420
Leu Pro Val Val Gly Gly Val Leu Gly Val Thr Gln Ala Val Val
                425                 430                 435
Leu Ser Ala Gly Phe Ser Ser Phe Tyr Leu Ala Asp Ile Asp Ser
                440                 445                 450
Gly Arg Asn Ile Phe Ile Val Gly Phe Ser Ile Phe Met Ala Leu
                455                 460                 465
Leu Leu Pro Arg Trp Phe Arg Glu Ala Pro Val Leu Phe Ser Thr
                470                 475                 480
Gly Trp Ser Pro Leu Asp Val Leu Leu His Ser Leu Leu Thr Gln
                485                 490                 495
Pro Ile Phe Leu Ala Gly Leu Ser Gly Phe Leu Leu Glu Asn Thr
                500                 505                 510
Ile Pro Gly Thr Gln Leu Glu Arg Gly Leu Gly Gln Gly Leu Pro
                515                 520                 525
Ser Pro Phe Thr Ala Gln Glu Ala Arg Met Pro Gln Lys Pro Arg
                530                 535                 540
Glu Lys Ala Ala Gln Val Tyr Arg Leu Pro Phe Pro Ile Gln Asn
                545                 550                 555
Leu Cys Pro Cys Ile Pro Gln Pro Leu His Cys Leu Cys Pro Leu
```

560 565 570
Pro Glu Asp Pro Gly Asp Glu Glu Gly Gly Ser Ser Glu Pro Glu
575 580 585
Glu Met Ala Asp Leu Leu Pro Gly Ser Gly Glu Pro Cys Pro Glu
590 595 600
Ser Ser Arg Glu Gly Phe Arg Ser Gln Lys
605 610

<210> 4
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2614283CD1

<400> 4
Met Glu Ala Lys Glu Lys Gln His Leu Leu Asp Ala Arg Pro Ala
1 5 10 15
Ile Arg Ser Tyr Thr Gly Ser Leu Trp Gln Glu Gly Ala Gly Trp
20 25 30
Ile Pro Leu Pro Arg Pro Gly Leu Asp Leu Gln Ala Ile Glu Leu
35 40 45
Ala Ala Gln Ser Asn His His Cys His Ala Gln Lys Gly Pro Asp
50 55 60
Ser His Cys Asp Pro Lys Lys Gly Lys Ala Gln Arg Gln Leu Tyr
65 70 75
Val Ala Ser Ala Ile Cys Leu Leu Phe Met Ile Gly Glu Val Val
80 85 90
Gly Gly Tyr Leu Ala His Ser Leu Ala Val Met Thr Asp Ala Ala
95 100 105
His Leu Leu Thr Asp Phe Ala Ser Met Leu Ile Ser Leu Phe Ser
110 115 120
Leu Trp Met Ser Ser Arg Pro Ala Thr Lys Thr Met Asn Phe Gly
125 130 135
Trp Gln Arg Ala Glu Ile Leu Gly Ala Leu Val Ser Val Leu Ser
140 145 150
Ile Trp Val Val Thr Gly Val Leu Val Tyr Leu Ala Val Glu Arg
155 160 165
Leu Ile Ser Gly Asp Tyr Glu Ile Asp Gly Gly Thr Met Leu Ile
170 175 180
Thr Ser Gly Cys Ala Val Ala Val Asn Ile Ile Met Gly Leu Thr
185 190 195
Leu His Gln Ser Gly His Gly His Ser His Gly Thr Thr Asn Gln
200 205 210
Gln Glu Glu Asn Pro Ser Val Arg Ala Ala Phe Ile His Val Ile
215 220 225
Gly Asp Phe Met Gln Ser Met Gly Val Leu Val Ala Ala Tyr Ile
230 235 240
Leu Tyr Phe Lys Pro Glu Tyr Lys Tyr Val Asp Pro Ile Cys Thr
245 250 255
Phe Val Phe Ser Ile Leu Val Leu Gly Thr Thr Leu Thr Ile Leu
260 265 270
Arg Asp Val Ile Leu Val Leu Met Glu Gly Thr Pro Lys Gly Val
275 280 285

Asp Phe Thr Ala Val Arg Asp Leu Leu Leu Ser Val Glu Gly Val
290 295 300
Glu Ala Leu His Ser Leu His Ile Trp Ala Leu Thr Val Ala Gln
305 310 315
Pro Val Leu Ser Val His Ile Ala Ile Ala Gln Asn Thr Asp Ala
320 325 330
Gln Ala Val Leu Lys Thr Ala Ser Ser Arg Leu Gln Gly Lys Phe
335 340 345
His Phe His Thr Val Thr Ile Gln Ile Glu Asp Tyr Ser Glu Asp
350 355 360
Met Lys Asp Cys Gln Ala Cys Gln Gly Pro Ser Asp
365 370

<210> 5
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2667691CD1

<400> 5
Met Thr Gln Gly Lys Lys Lys Lys Arg Ala Ala Asn Arg Ser Ile
1 5 10 15
Met Leu Ala Lys Lys Ile Ile Ile Lys Asp Gly Gly Thr Pro Gln
20 25 30
Gly Ile Gly Ser Pro Ser Val Tyr His Ala Val Ile Val Ile Phe
35 40 45
Leu Glu Phe Phe Ala Trp Gly Leu Leu Thr Ala Pro Thr Leu Val
50 55 60
Val Leu His Glu Thr Phe Pro Lys His Thr Phe Leu Met Asn Gly
65 70 75
Leu Ile Gln Gly Val Lys Gly Leu Leu Ser Phe Leu Ser Ala Pro
80 85 90
Leu Ile Gly Ala Leu Ser Asp Val Trp Gly Arg Lys Ser Phe Leu
95 100 105
Leu Leu Thr Val Phe Phe Thr Cys Ala Pro Ile Pro Leu Met Lys
110 115 120
Ile Ser Pro Trp Trp Tyr Phe Ala Val Ile Ser Val Ser Gly Val
125 130 135
Phe Ala Val Thr Phe Ser Val Val Phe Ala Tyr Val Ala Asp Ile
140 145 150
Thr Gln Glu His Glu Arg Ser Met Ala Tyr Gly Leu Val Ser Ala
155 160 165
Thr Phe Ala Ala Ser Leu Val Thr Ser Pro Ala Ile Gly Ala Tyr
170 175 180
Leu Gly Arg Val Tyr Gly Asp Ser Leu Val Val Val Leu Ala Thr
185 190 195
Ala Ile Ala Leu Leu Asp Ile Cys Phe Ile Leu Val Ala Val Pro
200 205 210
Glu Ser Leu Pro Glu Lys Met Arg Pro Ala Ser Trp Gly Ala Pro
215 220 225
Ile Ser Trp Glu Gln Ala Asp Pro Phe Ala Ser Leu Lys Lys Val
230 235 240
Gly Gln Asp Ser Ile Val Leu Leu Ile Cys Ile Thr Val Phe Leu

245 250 255
Ser Tyr Leu Pro Glu Ala Gly Gln Tyr Ser Ser Phe Phe Leu Tyr
260 265 270
Leu Arg Gln Ile Met Lys Phe Ser Pro Glu Ser Val Ala Ala Phe
275 280 285
Ile Ala Val Leu Gly Ile Leu Ser Ile Ile Ala Gln Thr Ile Val
290 295 300
Leu Ser Leu Leu Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu
305 310 315
Leu Gly Leu Gly Phe Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe
320 325 330
Gly Ser Glu Pro Trp Met Met Trp Ala Ala Gly Ala Val Ala Ala
335 340 345
Met Ser Ser Ile Thr Phe Pro Ala Val Ser Ala Leu Val Ser Arg
350 355 360
Thr Ala Asp Ala Asp Gln Gln Gly Val Val Gln Gly Met Ile Thr
365 370 375
Gly Ile Arg Gly Leu Cys Asn Gly Leu Gly Pro Ala Leu Tyr Gly
380 385 390
Phe Ile Phe Tyr Ile Phe His Val Glu Leu Lys Glu Leu Pro Ile
395 400 405
Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro Gln His His Phe
410 415 420
Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu Phe Gly Ala
425 430 435
Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile Pro Glu
440 445 450
His Thr Asn Leu Ser Leu Arg Ser Ser Ser Trp Arg Lys His Cys
455 460 465
Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala
470 475 480
Lys Glu Pro Leu Leu Gln Asp Thr Asn Val
485 490

<210> 6
<211> 377
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3211415CD1

<400> 6
Met Leu Pro Leu Ser Ile Lys Asp Asp Glu Tyr Lys Pro Pro Lys
1 5 10 15
Phe Asn Leu Phe Gly Lys Ile Ser Gly Trp Phe Arg Ser Ile Leu
20 25 30
Ser Asp Lys Thr Ser Arg Asn Leu Phe Phe Phe Leu Cys Leu Asn
35 40 45
Leu Ser Phe Ala Phe Val Glu Leu Leu Tyr Gly Ile Trp Ser Asn
50 55 60
Cys Leu Gly Leu Ile Ser Asp Ser Phe His Met Phe Phe Asp Ser
65 70 75
Thr Ala Ile Leu Ala Gly Leu Ala Ala Ser Val Ile Ser Lys Trp
80 85 90

```
Arg Asp Asn Asp Ala Phe Ser Tyr Gly Tyr Val Arg Ala Glu Val
                95                  100                 105
Leu Ala Gly Phe Val Asn Gly Leu Phe Leu Ile Phe Thr Ala Phe
                110                 115                 120
Phe Ile Phe Ser Glu Gly Val Glu Arg Ala Leu Ala Pro Pro Asp
                125                 130                 135
Val His His Glu Arg Leu Leu Leu Val Ser Ile Leu Gly Phe Val
                140                 145                 150
Val Asn Leu Ile Gly Ile Phe Val Phe Lys His Gly Gly His Gly
                155                 160                 165
His Ser His Gly Ser Gly Gly His Gly His Ser His Ser Leu Phe
                170                 175                 180
Asn Gly Ala Leu Asp Gln Ala His Gly His Val Asp His Cys His
                185                 190                 195
Ser His Glu Val Lys His Gly Ala Ala His Ser His Asp His Ala
                200                 205                 210
His Gly His Gly His Phe His Ser His Asp Gly Pro Ser Leu Lys
                215                 220                 225
Glu Thr Thr Gly Pro Ser Arg Gln Ile Leu Gln Gly Val Phe Leu
                230                 235                 240
His Ile Leu Ala Asp Thr Leu Gly Ser Ile Gly Val Ile Ala Ser
                245                 250                 255
Ala Ile Met Met Gln Asn Phe Gly Leu Met Ile Ala Asp Pro Ile
                260                 265                 270
Cys Ser Ile Leu Ile Ala Ile Leu Ile Val Val Ser Val Ile Pro
                275                 280                 285
Leu Leu Arg Glu Ser Val Gly Ile Leu Met Gln Arg Thr Pro Pro
                290                 295                 300
Leu Leu Glu Asn Ser Leu Pro Gln Cys Tyr Gln Arg Val Gln Gln
                305                 310                 315
Leu Gln Gly Val Tyr Ser Leu Gln Glu Gln His Phe Trp Thr Leu
                320                 325                 330
Cys Ser Asp Val Tyr Val Gly Thr Leu Lys Leu Ile Val Ala Pro
                335                 340                 345
Asp Ala Asp Ala Arg Trp Ile Leu Ser Gln Thr His Asn Ile Phe
                350                 355                 360
Thr Gln Ala Gly Val Arg Gln Leu Tyr Val Gln Ile Asp Phe Ala
                365                 370                 375
Ala Met


<210> 7
<211> 340
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4739923CD1

<400> 7
Met Ala Asp Thr Ala Thr Thr Ala Ser Ala Ala Ala Ala Ser Ala
  1               5                   10                  15
Ala Ser Ala Ser Ser Asp Ala Pro Pro Phe Gln Leu Gly Lys Pro
                20                  25                  30
Arg Phe Gln Gln Thr Ser Phe Tyr Gly Arg Phe Arg His Phe Leu
```

35 40 45
Asp Ile Ile Asp Pro Arg Thr Leu Phe Val Thr Glu Arg Arg Leu
50 55 60
Arg Glu Ala Val Gln Leu Leu Glu Asp Tyr Lys His Gly Thr Leu
65 70 75
Arg Pro Gly Val Thr Asn Glu Gln Leu Trp Ser Ala Gln Lys Ile
80 85 90
Lys Gln Ala Ile Leu His Pro Asp Thr Asn Glu Lys Ile Phe Met
95 100 105
Pro Phe Arg Met Pro Gly Tyr Ile Pro Phe Gly Thr Pro Ile Val
110 115 120
Val Gly Leu Leu Leu Pro Asn Gln Thr Leu Ala Ser Thr Val Phe
125 130 135
Trp Gln Trp Leu Asn Gln Ser His Asn Ala Cys Val Asn Tyr Ala
140 145 150
Asn Arg Asn Ala Thr Lys Pro Ser Pro Ala Ser Lys Phe Ile Gln
155 160 165
Gly Tyr Leu Gly Ala Val Ile Ser Ala Val Ser Ile Ala Val Gly
170 175 180
Leu Asn Val Leu Val Gln Lys Ala Asn Lys Leu Thr Pro Ala Thr
185 190 195
Arg Leu Leu Ile Gln Arg Phe Val Pro Phe Pro Ala Val Ala Ser
200 205 210
Ala Asn Ile Cys Asn Val Val Leu Met Arg Tyr Gly Glu Leu Glu
215 220 225
Glu Gly Ile Asp Val Leu Asp Ser Asp Gly Asn Leu Val Gly Ser
230 235 240
Ser Lys Ile Ala Ala Arg His Ala Leu Leu Glu Thr Ala Leu Thr
245 250 255
Arg Val Val Leu Pro Met Pro Ile Leu Val Leu Pro Pro Ile Val
260 265 270
Met Ser Met Leu Glu Lys Thr Ala Leu Leu Gln Ala Arg Pro Arg
275 280 285
Leu Leu Leu Pro Val Gln Ser Leu Val Cys Leu Ala Ala Phe Gly
290 295 300
Leu Ala Leu Pro Leu Ala Ile Ser Leu Phe Pro Gln Met Ser Glu
305 310 315
Ile Glu Thr Ser Gln Leu Glu Pro Glu Ile Ala Gln Ala Thr Ser
320 325 330
Ser Arg Thr Val Val Tyr Asn Lys Gly Leu
335 340

<210> 8
<211> 1274
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55030459CD1

<400> 8
Met Ala Arg Gln Pro Glu Glu Glu Glu Thr Ala Val Ala Arg Ala
1 5 10 15
Arg Arg Pro Pro Leu Trp Leu Leu Cys Leu Val Ala Cys Trp Leu
20 25 30

Leu Gly Ala Gly Ala Glu Ala Asp Phe Ser Ile Leu Asp Glu Ala
35 40 45
Gln Val Leu Ala Ser Gln Met Arg Arg Leu Ala Ala Glu Glu Leu
50 55 60
Gly Val Val Thr Met Gln Arg Ile Phe Asn Ser Phe Val Tyr Thr
65 70 75
Glu Lys Ile Ser Asn Gly Glu Ser Glu Val Gln Gln Leu Ala Lys
80 85 90
Lys Ile Arg Glu Lys Phe Asn Arg Tyr Leu Asp Val Val Asn Arg
95 100 105
Asn Lys Gln Val Val Glu Ala Ser Tyr Thr Ala His Leu Thr Ser
110 115 120
Pro Leu Thr Ala Ile Gln Asp Cys Cys Thr Ile Pro Pro Ser Met
125 130 135
Met Glu Phe Asp Gly Asn Phe Asn Thr Asn Val Ser Arg Thr Ile
140 145 150
Ser Cys Asp Arg Leu Ser Thr Thr Val Asn Ser Arg Ala Phe Asn
155 160 165
Pro Gly Arg Asp Leu Asn Ser Val Leu Ala Asp Asn Leu Lys Ser
170 175 180
Asn Pro Gly Ile Lys Trp Gln Tyr Phe Ser Ser Glu Glu Gly Ile
185 190 195
Phe Thr Val Phe Pro Ala His Lys Phe Arg Cys Lys Gly Ser Tyr
200 205 210
Glu His Arg Ser Arg Pro Ile Tyr Val Ser Thr Val Arg Pro Gln
215 220 225
Ser Lys His Ile Val Val Ile Leu Asp His Gly Ala Ser Val Thr
230 235 240
Asp Thr Gln Leu Gln Ile Ala Lys Asp Ala Ala Gln Val Ile Leu
245 250 255
Ser Ala Ile Asp Glu His Asp Lys Ile Ser Val Leu Thr Val Ala
260 265 270
Asp Thr Val Arg Thr Cys Ser Leu Asp Gln Cys Tyr Lys Thr Phe
275 280 285
Leu Ser Pro Ala Thr Ser Glu Thr Lys Arg Lys Met Ser Thr Phe
290 295 300
Val Ser Ser Val Lys Ser Ser Asp Ser Pro Thr Gln His Ala Val
305 310 315
Gly Phe Gln Lys Ala Phe Gln Leu Ile Arg Ser Thr Asn Asn Asn
320 325 330
Thr Lys Phe Gln Ala Asn Thr Asp Met Val Ile Ile Tyr Leu Ser
335 340 345
Ala Gly Ile Thr Ser Lys Asp Ser Ser Glu Glu Asp Lys Lys Ala
350 355 360
Thr Leu Gln Val Ile Asn Glu Glu Asn Ser Phe Leu Asn Asn Ser
365 370 375
Val Met Ile Leu Thr Tyr Ala Leu Met Asn Asp Gly Val Thr Gly
380 385 390
Leu Lys Glu Leu Ala Phe Leu Arg Asp Leu Ala Glu Gln Asn Ser
395 400 405
Gly Lys Tyr Gly Val Pro Asp Arg Thr Ala Leu Pro Val Ile Lys
410 415 420
Gly Ser Met Met Val Leu Asn Gln Leu Ser Asn Leu Glu Thr Thr
425 430 435
Val Gly Arg Phe Tyr Thr Asn Leu Pro Asn Arg Met Ile Asp Glu
440 445 450

Ala Val Phe Ser Leu Pro Phe Ser Asp Glu Met Gly Asp Gly Leu
455 460 465
Ile Met Thr Val Ser Lys Pro Cys Tyr Phe Gly Asn Leu Leu Leu
470 475 480
Gly Ile Val Gly Val Asp Val Asn Leu Ala Tyr Ile Leu Glu Asp
485 490 495
Val Thr Tyr Tyr Gln Asp Ser Leu Ala Ser Tyr Thr Phe Leu Ile
500 505 510
Asp Asp Lys Gly Tyr Thr Leu Met His Pro Ser Leu Thr Arg Pro
515 520 525
Tyr Leu Leu Ser Glu Pro Pro Leu His Thr Asp Ile Ile His Tyr
530 535 540
Glu Asn Ile Pro Lys Phe Glu Leu Val Arg Gln Asn Ile Leu Ser
545 550 555
Leu Pro Leu Gly Ser Gln Ile Ile Ala Val Pro Val Asn Ser Ser
560 565 570
Leu Ser Trp His Ile Asn Lys Leu Arg Glu Thr Gly Lys Glu Ala
575 580 585
Tyr Asn Val Ser Tyr Ala Trp Lys Met Val Gln Asp Thr Ser Phe
590 595 600
Ile Leu Cys Ile Val Val Ile Gln Pro Glu Ile Pro Val Lys Gln
605 610 615
Leu Lys Asn Leu Asn Thr Val Pro Ser Ser Lys Leu Leu Tyr His
620 625 630
Arg Leu Asp Leu Leu Gly Gln Pro Ser Ala Cys Leu His Phe Lys
635 640 645
Gln Leu Ala Thr Leu Glu Ser Pro Thr Ile Met Leu Ser Ala Gly
650 655 660
Ser Phe Ser Ser Pro Tyr Glu His Leu Ser Gln Pro Glu Thr Lys
665 670 675
Arg Met Val Glu His Tyr Thr Ala Tyr Leu Ser Asp Asn Thr Arg
680 685 690
Leu Ile Ala Asn Pro Gly Leu Lys Phe Ser Val Arg Asn Glu Val
695 700 705
Met Ala Thr Ser His Val Thr Asp Glu Trp Met Thr Gln Met Glu
710 715 720
Met Ser Ser Leu Asn Thr Tyr Ile Val Arg Arg Tyr Ile Ala Thr
725 730 735
Pro Asn Gly Val Leu Arg Ile Tyr Pro Gly Ser Leu Met Asp Lys
740 745 750
Ala Phe Asp Pro Thr Arg Arg Gln Trp Tyr Leu His Ala Val Ala
755 760 765
Asn Pro Gly Leu Ile Ser Leu Thr Gly Pro Tyr Leu Asp Val Gly
770 775 780
Gly Ala Gly Tyr Val Val Thr Ile Ser His Thr Ile His Ser Ser
785 790 795
Ser Thr Gln Leu Ser Ser Gly His Thr Val Ala Val Met Gly Ile
800 805 810
Asp Phe Thr Leu Arg Tyr Phe Tyr Lys Val Leu Met Asp Leu Leu
815 820 825
Pro Val Cys Asn Gln Asp Gly Gly Asn Lys Ile Arg Cys Phe Ile
830 835 840
Met Glu Asp Arg Gly Tyr Leu Val Ala His Pro Thr Leu Ile Asp
845 850 855
Pro Lys Gly His Ala Pro Val Glu Gln Gln His Ile Thr His Lys
860 865 870

```
Glu Pro Leu Val Ala Asn Asp Ile Leu Asn His Pro Asn Phe Val
                875                 880                 885
Lys Lys Asn Leu Cys Asn Ser Phe Ser Asp Arg Thr Val Gln Arg
                890                 895                 900
Phe Tyr Lys Phe Asn Thr Ser Leu Ala Gly Asp Leu Thr Asn Leu
                905                 910                 915
Val His Gly Ser His Cys Ser Lys Tyr Arg Leu Ala Arg Ile Pro
                920                 925                 930
Gly Thr Asn Ala Phe Val Gly Ile Val Asn Glu Thr Cys Asp Ser
                935                 940                 945
Leu Ala Phe Cys Ala Cys Ser Met Val Asp Arg Leu Cys Leu Asn
                950                 955                 960
Cys His Arg Met Glu Gln Asn Glu Cys Glu Cys Pro Cys Glu Cys
                965                 970                 975
Pro Leu Glu Val Asn Glu Cys Thr Gly Asn Leu Thr Asn Ala Glu
                980                 985                 990
Asn Arg Asn Pro Ser Cys Glu Val His Gln Glu Pro Val Thr Tyr
                995                 1000                1005
Thr Ala Ile Asp Pro Gly Leu Gln Asp Ala Leu His Gln Cys Val
                1010                1015                1020
Asn Ser Arg Cys Ser Gln Arg Leu Glu Ser Gly Asp Cys Phe Gly
                1025                1030                1035
Val Leu Asp Cys Glu Trp Cys Met Val Asp Ser Asp Gly Lys Thr
                1040                1045                1050
His Leu Asp Lys Pro Tyr Cys Ala Pro Gln Lys Glu Cys Phe Gly
                1055                1060                1065
Gly Ile Val Gly Ala Lys Ser Pro Tyr Val Asp Asp Met Gly Ala
                1070                1075                1080
Ile Gly Asp Glu Val Ile Thr Leu Asn Met Ile Lys Ser Ala Pro
                1085                1090                1095
Val Gly Pro Val Ala Gly Gly Ile Met Gly Cys Ile Met Val Leu
                1100                1105                1110
Val Leu Ala Val Tyr Ala Tyr Arg His Gln Ile His Arg Arg Ser
                1115                1120                1125
His Gln His Met Ser Pro Leu Ala Ala Gln Glu Met Ser Val Arg
                1130                1135                1140
Met Ser Asn Leu Glu Asn Asp Arg Asp Glu Arg Asp Asp Asp Ser
                1145                1150                1155
His Glu Asp Arg Gly Ile Ile Ser Asn Thr Arg Phe Ile Ala Ala
                1160                1165                1170
Val Ile Glu Arg His Ala His Ser Pro Glu Arg Arg Arg Arg Tyr
                1175                1180                1185
Trp Gly Arg Ser Gly Thr Glu Ser Asp His Gly Tyr Ser Thr Met
                1190                1195                1200
Ser Pro Gln Glu Asp Ser Glu Asn Pro Pro Cys Asn Asn Asp Pro
                1205                1210                1215
Leu Ser Ala Gly Val Asp Val Gly Asn His Asp Glu Asp Leu Asp
                1220                1225                1230
Leu Asp Thr Pro Pro Gln Thr Ala Ala Leu Leu Ser His Lys Phe
                1235                1240                1245
His His Tyr Arg Ser His His Pro Thr Leu His His Ser His His
                1250                1255                1260
Leu Gln Ala Ala Val Thr Val His Thr Val Asp Ala Glu Cys
                1265                1270
```

<210> 9

<211> 595
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6113039CD1

<400> 9
Met Lys Phe Phe Ser Tyr Ile Leu Val Tyr Arg Arg Phe Leu Phe
1 5 10 15
Val Val Phe Thr Val Leu Val Leu Leu Pro Leu Pro Ile Val Leu
20 25 30
His Thr Lys Glu Ala Glu Cys Ala Tyr Thr Leu Phe Val Val Ala
35 40 45
Thr Phe Trp Leu Thr Glu Ala Leu Pro Leu Ser Val Thr Ala Leu
50 55 60
Leu Pro Ser Leu Met Leu Pro Met Phe Gly Ile Met Pro Ser Lys
65 70 75
Lys Val Ala Ser Ala Tyr Phe Lys Asp Phe His Leu Leu Leu Ile
80 85 90
Gly Val Ile Cys Leu Ala Thr Ser Ile Glu Lys Trp Asn Leu His
95 100 105
Lys Arg Ile Ala Leu Lys Met Val Met Met Val Gly Val Asn Pro
110 115 120
Ala Trp Leu Thr Leu Gly Phe Met Ser Ser Thr Ala Phe Leu Ser
125 130 135
Met Trp Leu Ser Asn Thr Ser Thr Ala Ala Met Val Met Pro Ile
140 145 150
Ala Glu Ala Val Val Gln Gln Ile Ile Asn Ala Glu Ala Glu Val
155 160 165
Glu Ala Thr Gln Met Thr Tyr Phe Asn Gly Ser Thr Asn His Gly
170 175 180
Leu Glu Ile Asp Glu Ser Val Asn Gly His Glu Ile Asn Glu Arg
185 190 195
Lys Glu Lys Thr Lys Pro Val Pro Gly Tyr Asn Asn Asp Thr Gly
200 205 210
Lys Ile Ser Ser Lys Val Glu Leu Glu Lys Asn Ser Gly Met Arg
215 220 225
Thr Lys Tyr Arg Thr Lys Lys Gly His Val Thr Arg Lys Leu Thr
230 235 240
Cys Leu Cys Ile Ala Tyr Ser Ser Thr Ile Gly Gly Leu Thr Thr
245 250 255
Ile Thr Gly Thr Ser Thr Asn Leu Ile Phe Ala Glu Tyr Phe Asn
260 265 270
Thr Arg Tyr Pro Asp Cys Arg Cys Leu Asn Phe Gly Ser Trp Phe
275 280 285
Thr Phe Ser Phe Pro Ala Ala Leu Ile Ile Leu Leu Leu Ser Trp
290 295 300
Ile Trp Leu Gln Trp Leu Phe Leu Gly Phe Asn Phe Lys Glu Met
305 310 315
Phe Lys Cys Gly Lys Thr Lys Thr Val Gln Gln Lys Ala Cys Ala
320 325 330
Glu Val Ile Lys Gln Glu Tyr Gln Lys Leu Gly Pro Ile Arg Tyr
335 340 345
Gln Glu Ile Val Thr Leu Val Leu Phe Ile Ile Met Ala Leu Leu

350 355 360
Trp Phe Ser Arg Asp Pro Gly Phe Val Pro Gly Trp Ser Ala Leu
365 370 375
Phe Ser Glu Tyr Pro Gly Phe Ala Thr Asp Ser Thr Val Ala Leu
380 385 390
Leu Ile Gly Leu Leu Phe Phe Leu Ile Pro Ala Lys Thr Leu Thr
395 400 405
Lys Thr Thr Pro Thr Gly Glu Ile Val Ala Phe Asp Tyr Ser Pro
410 415 420
Leu Ile Thr Trp Lys Glu Phe Gln Ser Phe Met Pro Trp Asp Ile
425 430 435
Ala Ile Leu Val Gly Gly Gly Phe Ala Leu Ala Asp Gly Cys Glu
440 445 450
Glu Ser Gly Leu Ser Lys Trp Ile Gly Asn Lys Leu Ser Pro Leu
455 460 465
Gly Ser Leu Pro Ala Trp Leu Ile Ile Leu Ile Ser Ser Leu Met
470 475 480
Val Thr Ser Leu Thr Glu Val Ala Ser Asn Pro Ala Thr Ile Thr
485 490 495
Leu Phe Leu Pro Ile Leu Ser Pro Leu Ala Glu Ala Ile His Val
500 505 510
Asn Pro Leu Tyr Ile Leu Ile Pro Ser Thr Leu Cys Thr Ser Phe
515 520 525
Ala Phe Leu Leu Pro Val Ala Asn Pro Pro Asn Ala Ile Val Phe
530 535 540
Ser Tyr Gly His Leu Lys Val Ile Asp Met Val Lys Ala Gly Leu
545 550 555
Gly Val Asn Ile Val Gly Val Ala Val Val Met Leu Gly Ile Cys
560 565 570
Thr Trp Ile Val Pro Met Phe Asp Leu Tyr Thr Tyr Pro Ser Trp
575 580 585
Ala Pro Ala Met Ser Asn Glu Thr Met Pro
590 595

<210> 10
<211> 475
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7101781CD1

<400> 10
Met Ser Pro Glu Val Thr Cys Pro Arg Arg Gly His Leu Pro Arg
1 5 10 15
Phe His Pro Arg Thr Trp Val Glu Pro Val Val Ala Ser Ser Gln
20 25 30
Val Ala Ala Ser Leu Tyr Asp Ala Gly Leu Leu Leu Val Val Lys
35 40 45
Ala Ser Tyr Gly Thr Gly Gly Ser Ser Asn His Ser Ala Ser Pro
50 55 60
Ser Pro Arg Gly Ala Leu Glu Asp Gln Gln Gln Arg Ala Ile Ser
65 70 75
Asn Phe Tyr Ile Ile Tyr Asn Leu Val Val Gly Leu Ser Pro Leu
80 85 90

```
Leu Ser Ala Tyr Gly Leu Gly Trp Leu Ser Asp Arg Tyr His Arg
                95                  100                 105
Lys Ile Ser Ile Cys Met Ser Leu Leu Gly Phe Leu Leu Ser Arg
                110                 115                 120
Leu Gly Leu Leu Leu Lys Val Leu Leu Asp Trp Pro Val Glu Val
                125                 130                 135
Leu Tyr Gly Ala Ala Ala Leu Asn Gly Leu Phe Gly Gly Phe Ser
                140                 145                 150
Ala Phe Trp Ser Gly Val Met Ala Leu Gly Ser Leu Gly Ser Ser
                155                 160                 165
Glu Gly Arg Arg Ser Val Arg Leu Ile Leu Ile Asp Leu Met Leu
                170                 175                 180
Gly Leu Ala Gly Phe Cys Gly Ser Met Ala Ser Gly His Leu Phe
                185                 190                 195
Lys Gln Met Ala Gly His Ser Gly Gln Gly Leu Ile Leu Thr Ala
                200                 205                 210
Cys Ser Val Ser Cys Ala Ser Phe Ala Leu Leu Tyr Ser Leu Leu
                215                 220                 225
Val Leu Lys Val Pro Glu Ser Val Ala Lys Pro Ser Gln Glu Leu
                230                 235                 240
Pro Ala Val Asp Thr Val Ser Gly Thr Val Gly Thr Tyr Arg Thr
                245                 250                 255
Leu Asp Pro Asp Gln Leu Asp Gln Gln Tyr Ala Val Gly His Pro
                260                 265                 270
Pro Ser Pro Gly Lys Ala Lys Pro His Lys Thr Thr Ile Ala Leu
                275                 280                 285
Leu Phe Val Gly Ala Ile Ile Tyr Asp Leu Ala Val Val Gly Thr
                290                 295                 300
Val Asp Val Ile Pro Leu Phe Val Leu Arg Glu Pro Leu Gly Trp
                305                 310                 315
Asn Gln Val Gln Val Gly Tyr Gly Met Ala Ala Gly Tyr Thr Ile
                320                 325                 330
Phe Ile Thr Ser Phe Leu Gly Val Leu Val Phe Ser Arg Cys Phe
                335                 340                 345
Arg Asp Thr Thr Met Ile Met Ile Gly Met Val Ser Phe Gly Ser
                350                 355                 360
Gly Ala Leu Leu Leu Ala Phe Val Lys Glu Thr Tyr Met Phe Tyr
                365                 370                 375
Ile Ala Arg Ala Val Met Leu Phe Ala Leu Ile Pro Val Thr Thr
                380                 385                 390
Ile Arg Ser Ala Met Ser Lys Leu Ile Lys Gly Ser Ser Tyr Gly
                395                 400                 405
Lys Val Phe Val Ile Leu Gln Leu Ser Leu Ala Leu Thr Gly Val
                410                 415                 420
Val Thr Ser Thr Leu Tyr Asn Lys Ile Tyr Gln Leu Thr Met Asp
                425                 430                 435
Met Phe Val Gly Ser Cys Phe Ala Leu Ser Ser Phe Leu Ser Phe
                440                 445                 450
Leu Ala Ile Ile Pro Ile Ser Ile Val Ala Tyr Lys Gln Val Pro
                455                 460                 465
Leu Ser Pro Tyr Gly Asp Ile Ile Glu Lys
                470                 475
```

<210> 11
<211> 927
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473036CD1

<400> 11
Met Gln Pro Ala Arg Gly Pro Leu Ala Ser Glu Pro Arg Thr Val
1 5 10 15
Leu Val Leu Arg Phe Cys Ala Ser Leu Met Glu Met Lys Leu Pro
20 25 30
Gly Gln Glu Gly Phe Glu Ala Ser Ser Ala Pro Arg Asn Ile Pro
35 40 45
Ser Gly Glu Leu Asp Ser Asn Pro Asp Pro Gly Thr Gly Pro Ser
50 55 60
Pro Asp Gly Pro Ser Asp Thr Glu Ser Lys Glu Leu Gly Val Pro
65 70 75
Lys Asp Pro Leu Leu Phe Ile Gln Leu Asn Glu Leu Leu Gly Trp
80 85 90
Pro Gln Ala Leu Glu Trp Arg Glu Thr Gly Arg Trp Val Leu Phe
95 100 105
Glu Glu Lys Leu Glu Val Ala Ala Gly Arg Trp Ser Ala Pro His
110 115 120
Val Pro Thr Leu Ala Leu Pro Ser Leu Gln Lys Leu Arg Ser Leu
125 130 135
Leu Ala Glu Gly Leu Val Leu Leu Asp Cys Pro Ala Gln Ser Leu
140 145 150
Leu Glu Leu Val Gly Ser Thr His Pro Arg Lys Ala Ser Asp Asn
155 160 165
Glu Glu Ala Pro Leu Arg Glu Gln Cys Gln Asn Pro Leu Arg Gln
170 175 180
Lys Leu Pro Pro Gly Ala Glu Ala Gly Thr Val Leu Ala Gly Glu
185 190 195
Leu Gly Phe Leu Ala Gln Pro Leu Gly Ala Phe Val Arg Leu Arg
200 205 210
Asn Pro Val Val Leu Gly Ser Leu Thr Glu Val Ser Leu Pro Ser
215 220 225
Arg Phe Phe Cys Leu Leu Leu Gly Pro Cys Met Leu Gly Lys Gly
230 235 240
Tyr His Glu Met Gly Arg Ala Ala Ala Val Leu Leu Ser Asp Pro
245 250 255
Gln Phe Gln Trp Ser Val Arg Arg Ala Ser Asn Leu His Asp Leu
260 265 270
Leu Ala Ala Leu Asp Ala Phe Leu Glu Glu Val Thr Val Leu Pro
275 280 285
Pro Gly Arg Trp Asp Pro Thr Ala Arg Ile Pro Pro Pro Lys Cys
290 295 300
Leu Pro Ser Gln His Lys Arg Leu Pro Ser Gln Gln Arg Glu Ile
305 310 315
Arg Gly Pro Ala Val Pro Arg Leu Thr Ser Ala Glu Asp Arg His
320 325 330
Arg His Gly Pro His Ala His Ser Pro Glu Leu Gln Arg Thr Gly
335 340 345
Ser Asp Phe Leu Asp Ala Leu His Leu Gln Cys Phe Ser Ala Val
350 355 360
Leu Tyr Ile Tyr Leu Ala Thr Val Thr Asn Ala Ile Thr Phe Gly

```
                365                 370                 375
Gly Leu Leu Gly Asp Ala Thr Asp Gly Ala Gln Gly Val Leu Glu
                380                 385                 390
Ser Phe Leu Gly Thr Ala Val Ala Gly Ala Ala Phe Cys Leu Met
                395                 400                 405
Ala Gly Gln Pro Leu Thr Ile Leu Ser Ser Thr Gly Pro Val Leu
                410                 415                 420
Val Phe Glu Arg Leu Leu Phe Ser Phe Ser Arg Asp Tyr Ser Leu
                425                 430                 435
Asp Tyr Leu Pro Phe Arg Leu Trp Val Gly Ile Trp Val Ala Thr
                440                 445                 450
Phe Cys Leu Val Leu Val Ala Thr Glu Ala Ser Val Leu Val Arg
                455                 460                 465
Tyr Phe Thr Arg Phe Thr Glu Glu Gly Phe Cys Ala Leu Ile Ser
                470                 475                 480
Leu Ile Phe Ile Tyr Asp Ala Val Gly Lys Met Leu Asn Leu Thr
                485                 490                 495
His Thr Tyr Pro Ile Gln Lys Pro Gly Ser Ser Ala Tyr Gly Cys
                500                 505                 510
Leu Cys Gln Tyr Pro Gly Pro Gly Gly Asn Glu Ser Gln Trp Ile
                515                 520                 525
Arg Thr Arg Pro Lys Asp Arg Asp Asp Ile Val Ser Met Asp Leu
                530                 535                 540
Gly Leu Ile Asn Ala Ser Leu Leu Pro Pro Pro Glu Cys Thr Arg
                545                 550                 555
Gln Gly Gly His Pro Arg Gly Pro Gly Cys His Thr Val Pro Asp
                560                 565                 570
Ile Ala Phe Phe Ser Leu Leu Leu Phe Leu Thr Ser Phe Phe Phe
                575                 580                 585
Ala Met Ala Leu Lys Cys Val Lys Thr Ser Arg Phe Phe Pro Ser
                590                 595                 600
Val Val Arg Lys Gly Leu Ser Asp Phe Ser Ser Val Leu Ala Ile
                605                 610                 615
Leu Leu Gly Cys Gly Leu Asp Ala Phe Leu Gly Leu Ala Thr Pro
                620                 625                 630
Lys Leu Met Val Pro Arg Glu Phe Lys Pro Thr Leu Pro Gly Arg
                635                 640                 645
Gly Trp Leu Val Ser Pro Phe Gly Ala Asn Pro Trp Trp Trp Ser
                650                 655                 660
Val Ala Ala Ala Leu Pro Ala Leu Leu Leu Ser Ile Leu Ile Phe
                665                 670                 675
Met Asp Gln Gln Ile Thr Ala Val Ile Leu Asn Arg Met Glu Tyr
                680                 685                 690
Arg Leu Gln Lys Gly Ala Gly Phe His Leu Asp Leu Phe Cys Val
                695                 700                 705
Ala Val Leu Met Leu Leu Thr Ser Ala Leu Gly Leu Pro Trp Tyr
                710                 715                 720
Val Ser Ala Thr Val Ile Ser Leu Ala His Met Asp Ser Leu Arg
                725                 730                 735
Arg Glu Ser Arg Ala Cys Ala Pro Gly Glu Arg Pro Asn Phe Leu
                740                 745                 750
Gly Ile Arg Glu Gln Arg Leu Thr Gly Leu Val Val Phe Ile Leu
                755                 760                 765
Thr Gly Ala Ser Ile Phe Leu Ala Pro Val Leu Lys Phe Ile Pro
                770                 775                 780
Met Pro Val Leu Tyr Gly Ile Phe Leu Tyr Met Gly Val Ala Ala
```

785 790 795
Leu Ser Ser Ile Gln Phe Thr Asn Arg Val Lys Leu Leu Leu Met
800 805 810
Pro Ala Lys His Gln Pro Asp Leu Leu Leu Leu Arg His Val Pro
815 820 825
Leu Thr Arg Val His Leu Phe Thr Ala Ile Gln Leu Ala Cys Leu
830 835 840
Gly Leu Leu Trp Ile Ile Lys Ser Thr Pro Ala Ala Ile Ile Phe
845 850 855
Pro Leu Met Leu Leu Gly Leu Val Gly Val Arg Lys Ala Leu Glu
860 865 870
Arg Val Phe Ser Pro Gln Glu Leu Leu Trp Leu Asp Glu Leu Met
875 880 885
Pro Glu Glu Glu Arg Ser Ile Pro Glu Lys Gly Leu Glu Pro Glu
890 895 900
His Ser Phe Ser Gly Ser Asp Ser Glu Asp Ser Glu Leu Met Tyr
905 910 915
Gln Pro Lys Ala Pro Glu Ile Asn Ile Ser Val Asn
920 925

<210> 12
<211> 516
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476943CD1

<400> 12
Met Pro Ser Gly Ser His Trp Thr Ala Asn Ser Ser Lys Ile Ile
1 5 10 15
Thr Trp Leu Leu Glu Gln Pro Gly Lys Glu Glu Lys Arg Lys Thr
20 25 30
Met Ala Lys Val Asn Arg Ala Arg Ser Thr Ser Pro Pro Asp Gly
35 40 45
Gly Trp Gly Trp Met Ile Val Ala Gly Cys Phe Leu Val Thr Ile
50 55 60
Cys Thr Arg Ala Val Thr Arg Cys Ile Ser Ile Phe Phe Val Glu
65 70 75
Phe Gln Thr Tyr Phe Thr Gln Asp Tyr Ala Gln Thr Ala Trp Ile
80 85 90
His Ser Ile Val Asp Cys Val Thr Met Leu Cys Ala Pro Leu Gly
95 100 105
Ser Val Val Ser Asn His Leu Ser Cys Gln Val Gly Ile Met Leu
110 115 120
Gly Gly Leu Leu Ala Ser Thr Gly Leu Ile Leu Ser Ser Phe Ala
125 130 135
Thr Ser Leu Lys His Leu Tyr Leu Thr Leu Gly Val Leu Thr Gly
140 145 150
Leu Gly Phe Ala Leu Cys Tyr Ser Pro Ala Ile Ala Met Val Gly
155 160 165
Lys Tyr Phe Ser Arg Arg Lys Ala Leu Ala Tyr Gly Ile Ala Met
170 175 180
Ser Gly Ser Gly Ile Gly Thr Phe Ile Leu Ala Pro Val Val Gln
185 190 195

```
Leu Leu Ile Glu Gln Phe Ser Trp Arg Gly Ala Leu Leu Ile Leu
                200                 205                 210
Gly Gly Phe Val Leu Asn Leu Cys Val Cys Gly Ala Leu Met Arg
                215                 220                 225
Pro Ile Thr Leu Lys Glu Asp His Thr Thr Pro Glu Gln Asn His
                230                 235                 240
Val Cys Arg Thr Gln Lys Glu Asp Ile Lys Arg Val Ser Pro Tyr
                245                 250                 255
Ser Ser Leu Thr Lys Glu Trp Ala Gln Thr Cys Leu Cys Cys Cys
                260                 265                 270
Leu Gln Gln Glu Tyr Ser Phe Leu Leu Met Ser Asp Phe Val Val
                275                 280                 285
Leu Ala Val Ser Val Leu Phe Met Ala Tyr Gly Cys Ser Pro Leu
                290                 295                 300
Phe Val Tyr Leu Val Pro Tyr Ala Leu Ser Val Gly Val Ser His
                305                 310                 315
Gln Gln Ala Ala Phe Leu Met Ser Ile Leu Gly Val Ile Asp Ile
                320                 325                 330
Ile Gly Asn Ile Thr Phe Gly Trp Leu Thr Asp Arg Arg Cys Leu
                335                 340                 345
Lys Asn Tyr Gln Tyr Val Cys Tyr Leu Phe Ala Val Gly Met Asp
                350                 355                 360
Gly Leu Cys Tyr Leu Cys Leu Pro Met Leu Gln Ser Leu Pro Leu
                365                 370                 375
Leu Val Pro Phe Ser Cys Thr Phe Gly Tyr Phe Asp Gly Ala Tyr
                380                 385                 390
Val Thr Leu Ile Pro Val Val Thr Thr Glu Ile Val Gly Thr Thr
                395                 400                 405
Ser Leu Ser Ser Ala Leu Gly Val Val Tyr Phe Leu His Ala Val
                410                 415                 420
Pro Tyr Leu Val Ser Pro Pro Ile Ala Gly Arg Leu Val Asp Thr
                425                 430                 435
Thr Gly Ser Tyr Thr Ala Ala Phe Leu Leu Cys Gly Phe Ser Met
                440                 445                 450
Ile Phe Ser Ser Val Leu Leu Gly Phe Ala Arg Leu Ile Lys Arg
                455                 460                 465
Met Arg Lys Thr Gln Leu Gln Phe Ile Ala Lys Glu Ser Asp Pro
                470                 475                 480
Lys Leu Gln Leu Trp Thr Asn Gly Ser Val Ala Tyr Ser Val Ala
                485                 490                 495
Arg Glu Leu Asp Gln Lys His Gly Glu Pro Val Ala Thr Ala Val
                500                 505                 510
Pro Gly Tyr Ser Leu Thr
                515
```

<210> 13
<211> 514
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8003355CD1

<400> 13

```
Met His Gly Gly Gln Gly Pro Leu Leu Leu Leu Leu Leu Leu Ala
```

```
1               5                   10                  15
Val Cys Leu Gly Ala Gln Gly Arg Asn Gln Glu Glu Arg Leu Leu
                20                  25                  30
Ala Asp Leu Met Gln Asn Tyr Asp Pro Asn Leu Arg Pro Ala Glu
                35                  40                  45
Arg Asp Ser Asp Val Val Asn Val Ser Leu Lys Leu Thr Leu Thr
                50                  55                  60
Asn Leu Ile Ser Leu Asn Glu Arg Glu Glu Ala Leu Thr Thr Asn
                65                  70                  75
Val Trp Ile Glu Val Gln Trp Cys Asp Tyr Arg Leu Arg Arg Asp
                80                  85                  90
Pro Arg Asp Tyr Glu Gly Leu Trp Val Leu Arg Val Pro Ser Thr
                95                  100                 105
Met Val Trp Arg Pro Asp Ile Val Leu Glu Asn Asn Ala Asp Gly
                110                 115                 120
Val Phe Glu Val Ala Leu Tyr Cys Asn Val Leu Val Ser Pro Asp
                125                 130                 135
Gly Cys Ile Tyr Trp Leu Pro Pro Ala Ile Phe Arg Ser Ala Cys
                140                 145                 150
Ser Ile Ser Val Thr Tyr Phe Pro Phe Asp Trp Gln Asn Cys Ser
                155                 160                 165
Leu Ile Phe Gln Ser Gln Thr Tyr Ser Thr Asn Glu Ile Asp Leu
                170                 175                 180
Gln Leu Ser Gln Glu Asp Gly Gln Thr Ile Glu Trp Ile Phe Ile
                185                 190                 195
Asp Pro Glu Ala Phe Thr Glu Asn Gly Glu Trp Ala Ile Gln His
                200                 205                 210
Arg Pro Ala Lys Met Leu Leu Asp Pro Ala Ala Pro Ala Gln Glu
                215                 220                 225
Ala Gly His Gln Lys Val Val Phe Tyr Leu Leu Ile Gln Arg Lys
                230                 235                 240
Pro Leu Phe Tyr Val Ile Asn Ile Ile Ala Pro Cys Val Leu Ile
                245                 250                 255
Ser Ser Val Ala Ile Leu Ile His Phe Leu Pro Ala Lys Ala Gly
                260                 265                 270
Gly Gln Lys Cys Thr Val Ala Ile Asn Val Leu Leu Ala Gln Thr
                275                 280                 285
Val Phe Leu Phe Leu Val Ala Lys Lys Val Pro Glu Thr Ser Gln
                290                 295                 300
Ala Val Pro Leu Ile Ser Lys Tyr Leu Thr Phe Leu Leu Val Val
                305                 310                 315
Thr Ile Leu Ile Val Val Asn Ala Val Val Val Leu Asn Val Ser
                320                 325                 330
Leu Arg Ser Pro His Thr His Ser Met Ala Arg Gly Val Phe Leu
                335                 340                 345
Arg Leu Leu Pro Gln Leu Leu Arg Met His Val Arg Pro Leu Ala
                350                 355                 360
Pro Ala Ala Val Gln Asp Thr Gln Ser Arg Leu Gln Asn Gly Ser
                365                 370                 375
Ser Gly Trp Ser Ile Thr Thr Gly Glu Glu Val Ala Leu Cys Leu
                380                 385                 390
Pro Arg Ser Glu Leu Leu Phe Gln Gln Trp Gln Arg Gln Gly Leu
                395                 400                 405
Val Ala Ala Ala Leu Glu Lys Leu Glu Lys Gly Pro Glu Leu Gly
                410                 415                 420
Leu Ser Gln Phe Cys Gly Ser Leu Lys Gln Ala Ala Pro Ala Ile
```

425 430 435
Gln Ala Cys Val Glu Ala Cys Asn Leu Ile Ala Cys Ala Arg His
440 445 450
Gln Gln Ser His Phe Asp Asn Gly Asn Glu Glu Trp Phe Leu Val
455 460 465
Gly Arg Val Leu Asp Arg Val Cys Phe Leu Ala Met Leu Ser Leu
470 475 480
Phe Ile Cys Gly Thr Ala Gly Ile Phe Leu Met Ala His Tyr Asn
485 490 495
Arg Val Pro Ala Leu Pro Phe Pro Gly Asp Pro Arg Pro Tyr Leu
500 505 510
Pro Ser Pro Asp

<210> 14
<211> 691
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3116448CD1

<400> 14
Met Glu Leu Arg Ser Thr Ala Ala Pro Arg Ala Glu Gly Tyr Ser
1 5 10 15
Asn Val Gly Phe Gln Asn Glu Glu Asn Phe Leu Glu Asn Glu Asn
20 25 30
Thr Ser Gly Asn Asn Ser Ile Arg Ser Arg Ala Val Gln Ser Arg
35 40 45
Glu His Thr Asn Thr Lys Gln Asp Glu Glu Gln Val Thr Val Glu
50 55 60
Gln Asp Ser Pro Arg Asn Arg Glu His Met Glu Asp Asp Asp Glu
65 70 75
Glu Met Gln Gln Lys Gly Cys Leu Glu Arg Arg Tyr Asp Thr Val
80 85 90
Cys Gly Phe Cys Arg Lys His Lys Thr Thr Leu Arg His Ile Ile
95 100 105
Trp Gly Ile Leu Leu Ala Gly Tyr Leu Val Met Val Ile Ser Ala
110 115 120
Cys Val Leu Asn Phe His Arg Ala Leu Pro Leu Phe Val Ile Thr
125 130 135
Val Ala Ala Ile Phe Phe Val Val Trp Asp His Leu Met Ala Lys
140 145 150
Tyr Glu His Arg Ile Asp Glu Met Leu Ser Pro Gly Arg Arg Leu
155 160 165
Leu Asn Ser His Trp Phe Trp Leu Lys Trp Val Ile Trp Ser Ser
170 175 180
Leu Val Leu Ala Val Ile Phe Trp Leu Ala Phe Asp Thr Ala Lys
185 190 195
Leu Gly Gln Gln Gln Leu Val Ser Phe Gly Gly Leu Ile Met Tyr
200 205 210
Ile Val Leu Leu Phe Leu Phe Ser Lys Tyr Pro Thr Arg Val Tyr
215 220 225
Trp Arg Pro Val Leu Trp Gly Ile Gly Leu Gln Phe Leu Leu Gly
230 235 240

Leu Leu Ile Leu Arg Thr Asp Pro Gly Phe Ile Ala Phe Asp Trp
245 250 255
Leu Gly Arg Gln Val Gln Thr Phe Leu Glu Tyr Thr Asp Ala Gly
260 265 270
Ala Ser Phe Gly Phe Gly Glu Lys Tyr Lys Asp His Phe Phe Gly
275 280 285
Phe Lys Val Leu Ala Ile Val Val Phe Phe Ser Thr Val Met Ser
290 295 300
Met Leu Tyr Tyr Leu Gly Leu Met Gln Trp Ile Ile Arg Lys Val
305 310 315
Gly Trp Ile Met Leu Val Thr Thr Gly Ser Ser Pro Ile Glu Ser
320 325 330
Val Val Ala Ser Gly Asn Ile Phe Val Gly Gln Thr Glu Ser Pro
335 340 345
Leu Leu Val Arg Pro Tyr Leu Pro Tyr Ile Thr Lys Ser Glu Leu
350 355 360
His Ala Ile Met Thr Ala Gly Phe Ser Thr Ile Ala Gly Ser Val
365 370 375
Leu Gly Ala Tyr Ile Ser Phe Gly Val Pro Ser Ser His Leu Leu
380 385 390
Thr Ala Ser Val Met Ser Ala Pro Ala Ser Leu Ala Ala Ala Lys
395 400 405
Leu Phe Trp Pro Glu Thr Glu Lys Pro Lys Ile Thr Leu Lys Asn
410 415 420
Ala Met Lys Met Glu Ser Gly Asp Ser Gly Asn Leu Leu Glu Ala
425 430 435
Ala Thr Gln Gly Ala Ser Ser Ser Ile Ser Leu Val Ala Asn Ile
440 445 450
Ala Val Asn Leu Ile Ala Phe Leu Ala Leu Leu Ser Phe Met Asn
455 460 465
Ser Ala Leu Ser Trp Phe Gly Asn Met Phe Asp Tyr Pro Gln Leu
470 475 480
Ser Phe Glu Leu Ile Cys Ser Tyr Ile Phe Met Pro Phe Ser Phe
485 490 495
Met Met Gly Val Glu Trp Gln Asp Ser Phe Met Val Ala Arg Leu
500 505 510
Ile Gly Tyr Lys Thr Phe Phe Asn Glu Phe Val Ala Tyr Glu His
515 520 525
Leu Ser Lys Trp Ile His Leu Arg Lys Glu Gly Gly Pro Lys Phe
530 535 540
Val Asn Gly Val Gln Gln Tyr Ile Ser Ile Arg Ser Glu Ile Ile
545 550 555
Ala Thr Tyr Ala Leu Cys Gly Phe Ala Asn Ile Gly Ser Leu Gly
560 565 570
Ile Val Ile Gly Gly Leu Thr Ser Met Ala Pro Ser Arg Lys Arg
575 580 585
Asp Ile Ala Ser Gly Ala Val Arg Ala Leu Ile Ala Gly Thr Val
590 595 600
Ala Cys Phe Met Thr Ala Cys Ile Ala Gly Ile Leu Ser Ser Thr
605 610 615
Pro Val Asp Ile Asn Cys His His Val Leu Glu Asn Ala Phe Asn
620 625 630
Ser Thr Phe Pro Gly Asn Thr Thr Lys Val Ile Ala Cys Cys Gln
635 640 645
Ser Leu Leu Ser Ser Thr Val Ala Lys Gly Pro Gly Glu Val Ile
650 655 660

Pro Gly Gly Asn His Ser Leu Tyr Ser Leu Lys Gly Cys Cys Thr
665 670 675
Leu Leu Asn Pro Ser Thr Phe Asn Cys Asn Gly Ile Ser Asn Thr
680 685 690
Phe

<210> 15
<211> 342
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 622868CD1

<400> 15
Met Lys Ser Arg Thr Trp Ala Ser Val His Leu His Ser Phe Phe
1 5 10 15
Ala Val Gly Thr Leu Leu Val Ala Leu Thr Gly Tyr Leu Val Arg
20 25 30
Thr Trp Trp Leu Tyr Gln Met Ile Leu Ser Thr Val Thr Val Pro
35 40 45
Phe Ile Leu Cys Cys Trp Val Leu Pro Glu Thr Pro Phe Trp Leu
50 55 60
Leu Ser Glu Gly Arg Tyr Glu Glu Ala Gln Lys Ile Val Asp Ile
65 70 75
Met Ala Lys Trp Asn Arg Ala Ser Ser Cys Lys Leu Ser Glu Leu
80 85 90
Leu Ser Leu Asp Leu Gln Gly Pro Val Ser Asn Ser Pro Thr Glu
95 100 105
Val Gln Lys His Asn Leu Ser Tyr Leu Phe Tyr Asn Trp Ser Ile
110 115 120
Thr Lys Arg Thr Leu Thr Val Trp Leu Ile Trp Phe Thr Gly Ser
125 130 135
Leu Gly Phe Tyr Ser Phe Ser Leu Asn Ser Val Asn Leu Gly Gly
140 145 150
Asn Glu Tyr Leu Asn Leu Phe Leu Leu Gly Val Val Glu Ile Pro
155 160 165
Ala Tyr Thr Phe Val Cys Ile Ala Thr Asp Lys Val Gly Arg Arg
170 175 180
Thr Val Leu Ala Tyr Ser Leu Phe Cys Ser Ala Leu Ala Cys Gly
185 190 195
Val Val Met Val Ile Pro Gln Lys His Tyr Ile Leu Gly Val Val
200 205 210
Thr Ala Met Val Gly Lys Phe Ala Ile Gly Ala Ala Phe Gly Leu
215 220 225
Ile Tyr Leu Tyr Thr Ala Glu Leu Tyr Pro Thr Ile Val Arg Ser
230 235 240
Leu Ala Val Gly Ser Gly Ser Met Val Cys Arg Leu Ala Ser Ile
245 250 255
Leu Ala Pro Phe Ser Val Asp Leu Ser Ser Ile Trp Ile Phe Ile
260 265 270
Pro Gln Leu Phe Val Gly Thr Met Ala Leu Leu Ser Gly Val Leu
275 280 285
Thr Leu Lys Leu Pro Glu Thr Leu Gly Lys Arg Leu Ala Thr Thr

```
                290                 295                 300
Trp Glu Glu Ala Ala Lys Leu Glu Ser Glu Asn Glu Ser Lys Ser
                305                 310                 315
Ser Lys Leu Leu Leu Thr Thr Asn Asn Ser Gly Leu Glu Lys Thr
                320                 325                 330
Glu Ala Ile Thr Pro Arg Asp Ser Gly Leu Gly Glu
                335                 340

<210> 16
<211> 791
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476494CD1

<400> 16
Met Gly His Phe Glu Lys Gly Gln His Ala Leu Leu Asn Glu Gly
  1               5                  10                  15
Glu Glu Asn Glu Met Glu Ile Phe Gly Tyr Arg Thr Gln Gly Cys
                 20                  25                  30
Arg Lys Ser Leu Cys Leu Ala Gly Ser Ile Phe Ser Phe Gly Ile
                 35                  40                  45
Leu Pro Leu Val Phe Tyr Trp Arg Pro Ala Trp His Val Trp Ala
                 50                  55                  60
His Cys Val Pro Cys Ser Leu Gln Glu Ala Asp Thr Val Leu Leu
                 65                  70                  75
Arg Thr Thr Val Arg Cys Ile Lys Val Gln Lys Ile Arg Tyr Val
                 80                  85                  90
Trp Asn Tyr Leu Glu Gly Gln Phe Gln Lys Ile Gly Ser Leu Glu
                 95                 100                 105
Asp Trp Leu Ser Ser Ala Lys Ile His Gln Lys Phe Gly Ser Gly
                110                 115                 120
Leu Thr Arg Glu Glu Gln Glu Ile Arg Arg Leu Met Cys Gly Pro
                125                 130                 135
Asn Thr Ile Asp Val Glu Val Thr Pro Ile Trp Lys Leu Leu Ile
                140                 145                 150
Lys Glu Val Leu Asn Pro Phe Tyr Ile Phe Gln Leu Phe Ser Val
                155                 160                 165
Cys Leu Trp Phe Ser Glu Asp Tyr Lys Glu Tyr Ala Phe Ala Ile
                170                 175                 180
Ile Ile Met Ser Ile Ile Ser Ile Ser Leu Thr Val Tyr Asp Leu
                185                 190                 195
Arg Glu Gln Ser Val Lys Leu His His Leu Val Glu Ser His Asn
                200                 205                 210
Ser Ile Thr Val Ser Val Cys Gly Arg Lys Ala Gly Val Gln Glu
                215                 220                 225
Leu Glu Ser Arg Val Leu Val Pro Gly Asp Leu Leu Ile Leu Thr
                230                 235                 240
Gly Asn Lys Val Leu Met Pro Cys Asp Ala Val Leu Ile Glu Gly
                245                 250                 255
Ser Cys Val Val Asp Glu Gly Met Leu Thr Gly Glu Ser Ile Pro
                260                 265                 270
Val Thr Lys Thr Pro Leu Pro Lys Met Asp Ser Ser Val Pro Trp
                275                 280                 285
```

Lys Thr Gln Ser Glu Ala Asp Tyr Lys Arg His Val Leu Phe Cys
290 295 300
Gly Thr Glu Val Ile Gln Ala Lys Ala Ala Cys Ser Gly Thr Val
305 310 315
Arg Ala Val Val Leu Gln Thr Gly Phe Asn Thr Ala Lys Gly Asp
320 325 330
Leu Val Arg Ser Ile Leu Tyr Pro Lys Pro Val Asn Phe Gln Leu
335 340 345
Tyr Arg Asp Ala Ile Arg Phe Leu Leu Cys Leu Val Gly Thr Ala
350 355 360
Thr Ile Gly Met Ile Tyr Thr Leu Cys Val Tyr Val Leu Ser Gly
365 370 375
Glu Pro Pro Glu Glu Val Val Arg Lys Ala Leu Asp Val Ile Thr
380 385 390
Ile Ala Val Pro Pro Ala Leu Pro Ala Ala Leu Thr Thr Gly Ile
395 400 405
Ile Tyr Ala Gln Arg Arg Leu Lys Lys Arg Gly Ile Phe Cys Ile
410 415 420
Ser Pro Gln Arg Ile Asn Val Cys Gly Gln Leu Asn Leu Val Cys
425 430 435
Phe Asp Lys Thr Gly Thr Leu Thr Arg Asp Gly Leu Asp Leu Trp
440 445 450
Gly Val Val Ser Cys Asp Arg Asn Gly Phe Gln Glu Val His Ser
455 460 465
Phe Ala Ser Gly Gln Ala Leu Pro Trp Gly Pro Leu Cys Ala Ala
470 475 480
Met Ala Ser Cys His Ser Leu Ile Leu Leu Asp Gly Thr Ile Gln
485 490 495
Gly Asp Pro Leu Asp Leu Lys Met Phe Glu Ala Thr Thr Trp Glu
500 505 510
Met Ala Phe Ser Gly Asp Asp Phe His Ile Lys Gly Val Pro Ala
515 520 525
His Ala Met Val Val Lys Pro Cys Arg Thr Ala Ser Gln Val Pro
530 535 540
Val Glu Gly Ile Ala Ile Leu His Gln Phe Pro Phe Ser Ser Ala
545 550 555
Leu Gln Arg Met Thr Val Ile Val Gln Glu Met Gly Gly Asp Arg
560 565 570
Leu Ala Phe Met Lys Gly Ala Pro Glu Arg Val Ala Ser Phe Cys
575 580 585
Gln Pro Glu Thr Val Pro Thr Ser Phe Val Ser Glu Leu Gln Ile
590 595 600
Tyr Thr Thr Gln Gly Phe Arg Val Ile Ala Leu Ala Tyr Lys Lys
605 610 615
Leu Glu Asn Asp His His Ala Thr Thr Leu Thr Arg Glu Thr Val
620 625 630
Glu Ser Asp Leu Ile Phe Leu Gly Leu Leu Ile Leu Glu Asn Arg
635 640 645
Leu Lys Glu Glu Thr Lys Pro Val Leu Glu Glu Leu Ile Ser Ala
650 655 660
Arg Ile Arg Thr Val Met Ile Thr Gly Asp Asn Leu Gln Thr Ala
665 670 675
Ile Thr Val Ala Arg Lys Ser Gly Met Val Ser Glu Ser Gln Lys
680 685 690
Val Ile Leu Ile Glu Ala Asn Glu Thr Thr Gly Ser Ser Ser Ala
695 700 705

Ser Ile Ser Trp Thr Leu Val Glu Glu Lys Lys His Ile Met Tyr
710 715 720
Gly Asn Gln Asp Asn Tyr Ile Asn Ile Arg Asp Glu Val Ser Asp
725 730 735
Lys Gly Arg Glu Gly Ser Tyr His Phe Ala Leu Thr Gly Lys Ser
740 745 750
Phe His Val Ile Ser Gln His Phe Ser Ser Leu Leu Pro Lys Ile
755 760 765
Leu Ile Asn Gly Thr Ile Phe Ala Arg Met Ser Pro Gly Gln Lys
770 775 780
Ser Ser Leu Val Glu Glu Phe Gln Lys Leu Glu
785 790

<210> 17
<211> 1108
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477260CD1

<400> 17
Met Val Thr Gly Gly Gln His His Pro Gly Ala Gly Leu Ser Phe
1 5 10 15
Thr Glu Leu Glu Asn Thr Phe Pro Leu Cys Leu Pro Pro Thr Pro
20 25 30
Phe Leu Leu Ala Leu Trp Ser Ser Cys Leu Pro Trp Asp Thr Gln
35 40 45
Gln Thr Cys Cys Pro Ser Phe Ala Gly Ser Pro Ala Ala Glu Gln
50 55 60
Leu Gln Asp Ile Leu Gly Glu Glu Asp Glu Ala Pro Asn Pro Thr
65 70 75
Leu Phe Thr Glu Met Asp Thr Leu Gln His Asp Gly Asp Gln Met
80 85 90
Glu Trp Lys Glu Ser Ala Arg Trp Ile Lys Phe Glu Glu Lys Val
95 100 105
Glu Glu Gly Gly Glu Arg Trp Ser Lys Pro His Val Ser Thr Leu
110 115 120
Ser Leu His Ser Leu Phe Glu Leu Arg Thr Cys Leu Gln Thr Gly
125 130 135
Thr Val Leu Leu Asp Leu Asp Ser Gly Ser Leu Pro Gln Ile Ile
140 145 150
Asp Asp Val Ile Glu Lys Gln Ile Glu Asp Gly Leu Leu Arg Pro
155 160 165
Glu Leu Arg Glu Arg Val Ser Tyr Val Leu Leu Arg Arg His Arg
170 175 180
His Gln Thr Lys Lys Pro Ile His Arg Ser Leu Ala Asp Ile Gly
185 190 195
Lys Ser Val Ser Thr Thr Asn Arg Ser Pro Ala Arg Ser Pro Gly
200 205 210
Ala Gly Pro Ser Leu His His Ser Thr Glu Asp Leu Arg Met Arg
215 220 225
Gln Ser Ala Asn Tyr Gly Arg Leu Cys His Ala Gln Ser Arg Ser
230 235 240
Met Asn Asp Ile Ser Leu Thr Pro Asn Thr Asp Gln Arg Lys Asn

245 250 255
Lys Phe Met Lys Lys Ile Pro Lys Asp Ser Glu Ala Ser Asn Val
260 265 270
Leu Val Gly Glu Val Asp Phe Leu Asp Gln Pro Phe Ile Ala Phe
275 280 285
Val Arg Leu Ile Gln Ser Ala Met Leu Gly Gly Val Thr Glu Val
290 295 300
Pro Val Pro Thr Arg Phe Leu Phe Ile Leu Leu Gly Pro Ser Gly
305 310 315
Arg Ala Lys Ser Tyr Asn Glu Ile Gly Arg Ala Ile Ala Thr Leu
320 325 330
Met Val Asp Asp Leu Phe Ser Asp Val Ala Tyr Lys Ala Arg Asn
335 340 345
Arg Glu Asp Leu Ile Ala Gly Ile Asp Glu Phe Leu Asp Glu Val
350 355 360
Ile Val Leu Pro Pro Gly Glu Trp Asp Pro Asn Ile Arg Ile Glu
365 370 375
Pro Pro Lys Lys Val Pro Ser Ala Asp Lys Arg Lys Ser Leu Phe
380 385 390
Ser Leu Ala Glu Leu Gly Gln Met Asn Gly Ser Val Gly Gly Gly
395 400 405
Gly Gly Ala Pro Gly Gly Gly Asn Gly Gly Gly Gly Gly Gly Gly
410 415 420
Ser Gly Gly Gly Ala Gly Ser Gly Gly Ala Gly Gly Thr Ser Ser
425 430 435
Gly Asp Asp Gly Glu Met Pro Ala Met His Glu Ile Gly Glu Glu
440 445 450
Leu Ile Trp Thr Gly Arg Phe Phe Gly Gly Leu Cys Leu Asp Ile
455 460 465
Lys Arg Lys Leu Pro Trp Phe Pro Ser Asp Phe Tyr Asp Gly Phe
470 475 480
His Ile Gln Ser Ile Ser Ala Ile Leu Phe Ile Tyr Leu Gly Cys
485 490 495
Ile Thr Asn Ala Ile Thr Phe Gly Gly Leu Leu Gly Asp Ala Thr
500 505 510
Asp Asn Tyr Gln Gly Val Met Glu Ser Phe Leu Gly Thr Ala Met
515 520 525
Ala Gly Ser Leu Phe Cys Leu Phe Ser Gly Gln Pro Leu Ile Ile
530 535 540
Leu Ser Ser Thr Gly Pro Ile Leu Ile Phe Glu Lys Leu Leu Phe
545 550 555
Asp Phe Ser Lys Gly Asn Gly Leu Asp Tyr Met Glu Phe Arg Leu
560 565 570
Trp Ile Gly Leu His Ser Ala Val Gln Cys Leu Ile Leu Val Ala
575 580 585
Thr Asp Ala Ser Phe Ile Ile Lys Tyr Ile Thr Arg Phe Thr Glu
590 595 600
Glu Gly Phe Ser Thr Leu Ile Ser Phe Ile Phe Ile Tyr Asp Ala
605 610 615
Ile Lys Lys Met Ile Gly Ala Phe Lys Tyr Tyr Pro Ile Asn Met
620 625 630
Asp Phe Lys Pro Asn Phe Ile Thr Thr Tyr Lys Cys Glu Cys Val
635 640 645
Ala Pro Asp Thr Gly Asp Leu Asn Thr Thr Val Phe Asn Ala Ser
650 655 660
Ala Pro Leu Ala Pro Asp Thr Asn Ala Ser Leu Tyr Asn Leu Leu

```
                        665                 670                 675
Asn Leu Thr Ala Leu Asp Trp Ser Leu Leu Ser Lys Lys Glu Cys
                680                 685                 690
Leu Ser Tyr Gly Gly Arg Leu Leu Gly Asn Ser Cys Lys Phe Ile
                695                 700                 705
Pro Asp Leu Ala Leu Met Ser Phe Ile Leu Phe Phe Gly Thr Tyr
                710                 715                 720
Ser Met Thr Leu Thr Leu Lys Lys Phe Lys Phe Ser Arg Tyr Phe
                725                 730                 735
Pro Thr Lys Val Arg Ala Leu Val Ala Asp Phe Ser Ile Val Phe
                740                 745                 750
Ser Ile Leu Met Phe Cys Gly Ile Asp Ala Cys Phe Gly Leu Glu
                755                 760                 765
Thr Pro Lys Leu His Val Pro Ser Val Ile Lys Pro Thr Arg Pro
                770                 775                 780
Asp Arg Gly Trp Phe Val Ala Pro Phe Gly Lys Asn Pro Trp Trp
                785                 790                 795
Val Tyr Pro Ala Ser Ile Leu Pro Ala Leu Leu Val Thr Ile Leu
                800                 805                 810
Ile Phe Met Asp Gln Gln Ile Thr Ala Val Ile Val Asn Arg Lys
                815                 820                 825
Glu Asn Lys Leu Lys Lys Ala Ala Gly Tyr His Leu Asp Leu Phe
                830                 835                 840
Trp Val Gly Ile Leu Met Ala Leu Cys Ser Phe Met Gly Leu Pro
                845                 850                 855
Trp Tyr Val Ala Ala Thr Val Ile Ser Ile Ala His Ile Asp Ser
                860                 865                 870
Leu Lys Met Glu Thr Glu Thr Ser Ala Pro Gly Glu Gln Pro Gln
                875                 880                 885
Phe Leu Gly Val Arg Glu Gln Arg Val Thr Gly Ile Ile Val Phe
                890                 895                 900
Ile Leu Thr Gly Ile Ser Val Phe Leu Ala Pro Ile Leu Lys Cys
                905                 910                 915
Ile Pro Leu Pro Val Leu Tyr Gly Val Phe Leu Tyr Met Gly Val
                920                 925                 930
Ala Ser Leu Asn Gly Ile Gln Phe Trp Glu Arg Cys Lys Leu Phe
                935                 940                 945
Leu Met Pro Ala Lys His Gln Pro Asp His Ala Phe Leu Arg His
                950                 955                 960
Val Pro Leu Arg Arg Ile His Leu Phe Thr Leu Val Gln Ile Leu
                965                 970                 975
Cys Leu Ala Val Leu Trp Ile Leu Lys Ser Thr Val Ala Ala Ile
                980                 985                 990
Ile Phe Pro Val Met Ile Leu Gly Leu Ile Ile Val Arg Arg Leu
                995                1000                1005
Leu Asp Phe Ile Phe Ser Gln His Asp Leu Ala Trp Ile Asp Asn
               1010                1015                1020
Ile Leu Pro Glu Lys Glu Lys Lys Glu Thr Asp Lys Lys Arg Lys
               1025                1030                1035
Arg Lys Lys Gly Ala His Glu Asp Cys Asp Glu Glu Glu Lys Asp
               1040                1045                1050
Leu Pro Val Gly Val Thr His Ser Asp Ser Ser Phe Ser Asp Thr
               1055                1060                1065
Glu Leu Asp Arg Ser Tyr Ser Arg Asn Pro Val Phe Met Val Pro
               1070                1075                1080
Gln Val Lys Ile Glu Met Glu Ser Asp Tyr Asp Phe Thr Asp Met
```

1085 1090 1095
Asp Lys Tyr Arg Arg Glu Thr Asp Ser Glu Thr Thr Leu
1100 1105

<210> 18
<211> 480
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1963058CD1

<400> 18
Met Gly Pro Gly Pro Pro Ala Ala Gly Ala Ala Pro Ser Pro Arg
1 5 10 15
Pro Leu Ser Leu Val Ala Arg Leu Ser Tyr Ala Val Gly His Phe
20 25 30
Leu Asn Asp Leu Cys Ala Ser Met Trp Phe Thr Tyr Leu Leu Leu
35 40 45
Tyr Leu His Ser Val Arg Ala Tyr Ser Ser Arg Gly Ala Gly Leu
50 55 60
Leu Leu Leu Leu Gly Gln Val Ala Asp Gly Leu Cys Thr Pro Leu
65 70 75
Val Gly Tyr Glu Ala Asp Arg Ala Ala Ser Cys Cys Ala Arg Tyr
80 85 90
Gly Pro Arg Lys Ala Trp His Leu Val Gly Thr Val Cys Val Leu
95 100 105
Leu Ser Phe Pro Phe Ile Phe Ser Pro Cys Leu Gly Cys Gly Ala
110 115 120
Ala Thr Pro Glu Trp Ala Ala Leu Leu Tyr Tyr Gly Pro Phe Ile
125 130 135
Val Ile Phe Gln Phe Gly Trp Ala Ser Thr Gln Ile Ser His Leu
140 145 150
Ser Leu Ile Pro Glu Leu Val Thr Asn Asp His Glu Lys Val Glu
155 160 165
Leu Thr Ala Leu Arg Tyr Ala Phe Thr Val Val Ala Asn Ile Thr
170 175 180
Val Tyr Gly Ala Ala Trp Leu Leu Leu His Leu Gln Gly Ser Ser
185 190 195
Arg Val Glu Pro Thr Gln Asp Ile Ser Ile Ser Asp Gln Leu Gly
200 205 210
Gly Gln Asp Val Pro Val Phe Arg Asn Leu Ser Leu Leu Val Val
215 220 225
Gly Val Gly Ala Val Phe Ser Leu Leu Phe His Leu Gly Thr Arg
230 235 240
Glu Arg Arg Arg Pro His Ala Glu Glu Pro Gly Glu His Thr Pro
245 250 255
Leu Leu Ala Pro Ala Thr Ala Gln Pro Leu Leu Leu Trp Lys His
260 265 270
Trp Leu Arg Glu Pro Ala Phe Tyr Gln Val Gly Ile Leu Tyr Met
275 280 285
Thr Thr Arg Leu Ile Val Asn Leu Ser Gln Thr Tyr Met Ala Met
290 295 300
Tyr Leu Thr Tyr Ser Leu His Leu Pro Lys Lys Phe Ile Ala Thr
305 310 315

Ile Pro Leu Val Met Tyr Leu Ser Gly Phe Leu Ser Ser Phe Leu
320 325 330
Met Lys Pro Ile Asn Lys Cys Ile Gly Arg Asn Met Thr Tyr Phe
335 340 345
Ser Gly Leu Leu Val Ile Leu Ala Phe Ala Ala Trp Val Ala Leu
350 355 360
Ala Glu Gly Leu Gly Val Ala Val Tyr Ala Ala Ala Val Leu Leu
365 370 375
Gly Ala Gly Cys Ala Thr Ile Leu Val Thr Ser Leu Ala Met Thr
380 385 390
Ala Asp Leu Ile Gly Pro His Thr Asn Ser Gly Ala Phe Val Tyr
395 400 405
Gly Ser Met Ser Phe Leu Asp Lys Val Ala Asn Gly Leu Ala Val
410 415 420
Met Ala Ile Gln Ser Leu His Pro Cys Pro Ser Glu Leu Cys Cys
425 430 435
Arg Ala Cys Val Ser Phe Tyr His Trp Ala Met Val Ala Val Thr
440 445 450
Gly Gly Val Gly Val Ala Ala Ala Leu Cys Leu Cys Ser Leu Leu
455 460 465
Leu Trp Pro Thr Arg Leu Arg Arg Trp Asp Arg Asp Ala Arg Pro
470 475 480

<210> 19
<211> 381
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2395967CD1

<400> 19
Met Ser Glu Phe Trp Leu Ile Ser Ala Pro Gly Asp Lys Glu Asn
1 5 10 15
Leu Gln Ala Leu Glu Arg Met Asn Thr Val Thr Ser Lys Ser Asn
20 25 30
Leu Ser Tyr Asn Thr Lys Phe Ala Ile Pro Asp Phe Lys Val Gly
35 40 45
Thr Leu Asp Ser Leu Val Gly Leu Ser Asp Glu Leu Gly Lys Leu
50 55 60
Asp Thr Phe Ala Glu Ser Leu Ile Arg Arg Met Ala Gln Ser Val
65 70 75
Val Glu Val Met Glu Asp Ser Lys Gly Lys Val Gln Glu His Leu
80 85 90
Leu Ala Asn Gly Val Asp Leu Thr Ser Phe Val Thr His Phe Glu
95 100 105
Trp Asp Met Ala Lys Tyr Pro Val Lys Gln Pro Leu Val Ser Val
110 115 120
Val Asp Thr Ile Ala Lys Gln Leu Ala Gln Ile Glu Met Asp Leu
125 130 135
Lys Ser Arg Thr Ala Ala Tyr Asn Thr Leu Lys Thr Asn Leu Glu
140 145 150
Asn Leu Glu Lys Lys Ser Met Gly Asn Leu Phe Thr Arg Thr Leu
155 160 165

Ser Asp Ile Val Ser Lys Glu Asp Phe Val Leu Asp Ser Glu Tyr
170 175 180
Leu Val Thr Leu Leu Val Ile Val Pro Lys Pro Asn Tyr Ser Gln
185 190 195
Trp Gln Lys Thr Tyr Glu Ser Leu Ser Asp Met Val Val Pro Arg
200 205 210
Ser Thr Lys Leu Ile Thr Glu Asp Lys Glu Gly Gly Leu Phe Thr
215 220 225
Val Thr Leu Phe Arg Lys Val Ile Glu Asp Phe Lys Thr Lys Ala
230 235 240
Lys Glu Asn Lys Phe Thr Val Arg Glu Phe Tyr Tyr Asp Glu Lys
245 250 255
Glu Ile Glu Arg Glu Arg Glu Glu Met Ala Arg Leu Leu Ser Asp
260 265 270
Lys Lys Gln Gln Tyr Gly Pro Leu Leu Arg Trp Leu Lys Val Asn
275 280 285
Phe Ser Glu Ala Phe Ile Ala Trp Ile His Ile Lys Ala Leu Arg
290 295 300
Val Phe Val Glu Ser Val Leu Arg Tyr Gly Leu Pro Val Asn Phe
305 310 315
Gln Ala Val Leu Leu Gln Pro His Lys Lys Ser Ser Thr Lys Arg
320 325 330
Leu Arg Glu Val Leu Asn Ser Val Phe Arg His Leu Asp Glu Val
335 340 345
Ala Ala Thr Ser Ile Leu Asp Ala Ser Val Glu Ile Pro Gly Leu
350 355 360
Gln Leu Asn Asn Gln Asp Tyr Phe Pro Tyr Val Tyr Phe His Ile
365 370 375
Asp Leu Ser Leu Leu Asp
380

<210> 20
<211> 484
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3586648CD1

<400> 20
Met Tyr Thr Ser His Glu Asp Ile Gly Tyr Asp Phe Glu Asp Gly
1 5 10 15
Pro Lys Asp Lys Lys Thr Leu Lys Pro His Pro Asn Ile Asp Gly
20 25 30
Gly Trp Ala Trp Met Met Val Leu Ser Ser Phe Phe Val His Ile
35 40 45
Leu Ile Met Gly Ser Gln Met Ala Leu Gly Val Leu Asn Val Glu
50 55 60
Trp Leu Glu Glu Phe His Gln Ser Arg Gly Leu Thr Ala Trp Val
65 70 75
Ser Ser Leu Ser Met Gly Ile Thr Leu Ile Val Gly Pro Phe Ile
80 85 90
Gly Leu Phe Ile Asn Thr Cys Gly Cys Arg Gln Thr Ala Ile Ile
95 100 105
Gly Gly Leu Val Asn Ser Leu Gly Trp Val Leu Ser Ala Tyr Ala

110 115 120
Ala Asn Val His Tyr Leu Phe Ile Thr Phe Gly Val Ala Ala Gly
125 130 135
Leu Gly Ser Gly Met Ala Tyr Leu Pro Ala Val Val Met Val Gly
140 145 150
Arg Tyr Phe Gln Lys Arg Arg Ala Leu Ala Gln Gly Leu Ser Thr
155 160 165
Thr Gly Thr Gly Phe Gly Thr Phe Leu Met Thr Val Leu Leu Lys
170 175 180
Tyr Leu Cys Ala Glu Tyr Gly Trp Arg Asn Ala Met Leu Ile Gln
185 190 195
Gly Ala Val Ser Leu Asn Leu Cys Val Cys Gly Ala Leu Met Arg
200 205 210
Pro Leu Ser Pro Gly Lys Asn Pro Asn Asp Pro Gly Glu Lys Asp
215 220 225
Val Arg Gly Leu Pro Ala His Ser Thr Glu Ser Val Lys Ser Thr
230 235 240
Gly Gln Gln Gly Arg Thr Glu Glu Lys Asp Gly Gly Leu Gly Asn
245 250 255
Glu Glu Thr Leu Cys Asp Leu Gln Ala Gln Glu Cys Pro Asp Gln
260 265 270
Ala Gly His Arg Lys Asn Met Cys Ala Leu Arg Ile Leu Lys Thr
275 280 285
Val Ser Trp Leu Thr Met Arg Val Arg Lys Gly Phe Glu Asp Trp
290 295 300
Tyr Ser Gly Tyr Phe Gly Thr Ala Ser Leu Phe Thr Asn Arg Met
305 310 315
Phe Val Ala Phe Ile Phe Trp Ala Leu Phe Ala Tyr Ser Ser Phe
320 325 330
Val Ile Pro Phe Ile His Leu Pro Glu Ile Val Asn Leu Tyr Asn
335 340 345
Leu Ser Glu Gln Asn Asp Val Phe Pro Leu Thr Ser Ile Ile Ala
350 355 360
Ile Val His Ile Phe Gly Lys Val Ile Leu Gly Val Ile Ala Asp
365 370 375
Leu Pro Cys Ile Ser Val Trp Asn Val Phe Leu Leu Ala Asn Phe
380 385 390
Thr Leu Val Leu Ser Ile Phe Ile Leu Pro Leu Met His Thr Tyr
395 400 405
Ala Gly Leu Ala Val Ile Cys Ala Leu Ile Gly Phe Ser Ser Gly
410 415 420
Tyr Phe Ser Leu Met Pro Val Val Thr Glu Asp Leu Val Gly Ile
425 430 435
Glu His Leu Ala Asn Ala Tyr Gly Ile Ile Ile Cys Ala Asn Gly
440 445 450
Ile Ser Ala Leu Leu Gly Pro Pro Phe Ala Gly Lys Leu Ser Glu
455 460 465
Val Leu Arg Ala Gln Ser Ala Cys Thr Tyr Gly Ala Leu Cys Tyr
470 475 480
Lys Val Pro Asp

<210> 21
<211> 736
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473396CD1

<400> 21
Met Gln Asn Ile Thr Lys Glu Phe Gly Thr Phe Lys Ala Asn Asp
1 5 10 15
Asn Ile Asn Leu Gln Val Lys Ala Gly Glu Ile His Ala Leu Leu
20 25 30
Gly Glu Asn Gly Ala Gly Lys Ser Thr Leu Met Asn Val Leu Ser
35 40 45
Gly Leu Leu Glu Pro Thr Ser Gly Lys Ile Leu Met Arg Gly Lys
50 55 60
Glu Val Gln Ile Thr Ser Pro Thr Lys Ala Asn Gln Leu Gly Ile
65 70 75
Gly Met Val His Gln His Phe Met Leu Val Asp Ala Phe Thr Val
80 85 90
Thr Glu Asn Ile Val Leu Gly Ser Glu Pro Ser Arg Ala Gly Met
95 100 105
Leu Asp His Lys Lys Ala Arg Lys Glu Ile Gln Lys Val Ser Glu
110 115 120
Gln Tyr Gly Leu Ser Val Asn Pro Asp Ala Tyr Val Arg Asp Ile
125 130 135
Ser Val Gly Met Glu Gln Arg Val Glu Ile Leu Lys Thr Leu Tyr
140 145 150
Arg Gly Ala Asp Val Leu Ile Phe Asp Glu Pro Thr Ala Val Leu
155 160 165
Thr Pro Gln Glu Ile Asp Glu Leu Ile Val Ile Met Lys Glu Leu
170 175 180
Val Lys Glu Gly Lys Ser Ile Ile Leu Ile Thr His Lys Leu Asp
185 190 195
Glu Ile Lys Ala Val Ala Asp Arg Cys Thr Val Ile Arg Arg Gly
200 205 210
Lys Gly Ile Gly Thr Val Asn Val Lys Asp Val Thr Ser Gln Gln
215 220 225
Leu Ala Asp Met Met Val Gly Arg Ala Val Ser Phe Lys Thr Met
230 235 240
Lys Lys Glu Ala Lys Pro Gln Glu Val Val Leu Ser Ile Glu Asn
245 250 255
Leu Val Val Lys Glu Asn Arg Gly Leu Glu Ala Val Lys Asn Leu
260 265 270
Asn Leu Glu Val Arg Ala Gly Glu Val Leu Gly Ile Ala Gly Ile
275 280 285
Asp Gly Asn Gly Gln Ser Glu Leu Ile Gln Ala Leu Thr Gly Leu
290 295 300
Arg Lys Ala Glu Ser Gly His Ile Lys Leu Lys Gly Glu Asp Ile
305 310 315
Thr Asn Lys Lys Pro Arg Lys Ile Thr Glu His Gly Val Gly His
320 325 330
Val Pro Glu Asp Arg His Lys Tyr Gly Leu Val Leu Asp Met Thr
335 340 345
Leu Ser Glu Asn Ile Ala Leu Gln Thr Tyr His Gln Lys Pro Tyr
350 355 360
Ser Lys Asn Gly Met Leu Asn Tyr Ser Val Ile Asn Glu His Ala
365 370 375
Arg Glu Leu Ile Glu Glu Tyr Asp Val Arg Thr Thr Asn Glu Leu

```
                380                 385                 390
Val Pro Ala Lys Ala Leu Ser Gly Gly Asn Gln Gln Lys Ala Ile
                395                 400                 405
Ile Ala Arg Ile Val Asp Arg Asp Pro Asp Leu Leu Ile Val Ala
                410                 415                 420
Asn Pro Thr Arg Gly Leu Asp Val Gly Glu Phe Val Ala Val Thr
                425                 430                 435
Gly Val Ser Gly Ser Gly Lys Ser Thr Leu Val Asn Ser Ile Leu
                440                 445                 450
Lys Lys Ser Leu Ala Gln Lys Leu Asn Lys Asn Ser Ala Lys Pro
                455                 460                 465
Gly Lys Phe Lys Thr Ile Ser Gly Tyr Glu Ser Ile Glu Lys Ile
                470                 475                 480
Ile Asp Ile Asp Gln Ser Pro Ile Gly Arg Thr Pro Arg Ser Asn
                485                 490                 495
Pro Ala Thr Tyr Thr Ser Val Phe Asp Asp Ile Arg Gly Leu Phe
                500                 505                 510
Ala Gln Thr Asn Glu Ala Lys Met Arg Gly Tyr Lys Lys Gly Arg
                515                 520                 525
Phe Ser Phe Asn Val Lys Gly Gly Arg Cys Glu Ala Cys Arg Gly
                530                 535                 540
Asp Gly Ile Ile Lys Ile Glu Met His Phe Leu Pro Asp Val Tyr
                545                 550                 555
Val Pro Cys Glu Val Cys His Gly Lys Arg Tyr Asn Ser Glu Thr
                560                 565                 570
Leu Glu Val His Tyr Lys Gly Lys Ser Ile Ala Asp Ile Leu Glu
                575                 580                 585
Met Thr Val Glu Asp Ala Val Glu Phe Phe Lys His Ile Pro Lys
                590                 595                 600
Ile His Arg Lys Leu Gln Thr Ile Val Asp Val Gly Leu Gly Tyr
                605                 610                 615
Val Thr Met Gly Gln Pro Ala Thr Thr Leu Ser Gly Gly Glu Ala
                620                 625                 630
Gln Arg Met Lys Leu Ala Ser Glu Leu His Lys Ile Ser Asn Gly
                635                 640                 645
Lys Asn Phe Tyr Ile Leu Asp Glu Pro Thr Thr Gly Leu His Ser
                650                 655                 660
Asp Asp Ile Ala Arg Leu Leu His Val Leu Gln Arg Leu Val Asp
                665                 670                 675
Ala Gly Asn Thr Val Leu Val Ile Glu His Asn Leu Asp Val Ile
                680                 685                 690
Lys Thr Ala Asp Tyr Ile Ile Asp Leu Gly Pro Glu Gly Gly Glu
                695                 700                 705
Gly Gly Gly Thr Ile Leu Thr Thr Gly Thr Pro Glu Glu Ile Ile
                710                 715                 720
Asn Val Lys Glu Ser Tyr Thr Gly His Tyr Leu Lys Lys Ile Met
                725                 730                 735
Val
```

<210> 22
<211> 465
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7476283CD1

<400> 22
Met Gly Pro Leu Lys Ala Phe Leu Phe Ser Pro Phe Leu Leu Arg
1 5 10 15
Ser Gln Ser Arg Gly Val Arg Leu Val Phe Leu Leu Leu Thr Leu
20 25 30
His Leu Gly Asn Cys Val Asp Lys Ala Asp Asp Glu Asp Asp Glu
35 40 45
Asp Leu Lys Val Asn Lys Thr Trp Val Leu Ala Pro Lys Ile His
50 55 60
Glu Gly Asp Ile Thr Gln Ile Leu Asn Ser Leu Leu Gln Gly Tyr
65 70 75
Asp Asn Lys Leu Arg Pro Asp Ile Gly Val Arg Pro Thr Val Ile
80 85 90
Glu Thr Asp Val Tyr Val Asn Ser Ile Gly Pro Val Asp Pro Ile
95 100 105
Asn Met Glu Tyr Thr Ile Asp Ile Ile Phe Ala Gln Thr Trp Phe
110 115 120
Asp Ser Arg Leu Lys Phe Asn Ser Thr Met Lys Val Leu Met Leu
125 130 135
Asn Ser Asn Met Val Gly Lys Ile Trp Ile Pro Asp Thr Phe Phe
140 145 150
Arg Asn Ser Arg Lys Ser Asp Ala His Trp Ile Thr Thr Pro Asn
155 160 165
Arg Leu Leu Arg Ile Trp Asn Asp Gly Arg Val Leu Tyr Thr Leu
170 175 180
Arg Leu Thr Ile Asn Ala Glu Cys Tyr Leu Gln Leu His Asn Phe
185 190 195
Pro Met Asp Glu His Ser Cys Pro Leu Glu Phe Ser Ser Asp Gly
200 205 210
Tyr Pro Lys Asn Glu Ile Glu Tyr Lys Trp Lys Lys Pro Ser Val
215 220 225
Glu Val Ala Asp Pro Lys Tyr Trp Arg Leu Tyr Gln Phe Ala Phe
230 235 240
Val Gly Leu Arg Asn Ser Thr Glu Ile Thr His Thr Ile Ser Gly
245 250 255
Asp Tyr Val Ile Met Thr Ile Phe Phe Asp Leu Ser Arg Arg Met
260 265 270
Gly Tyr Phe Thr Ile Gln Thr Tyr Ile Pro Cys Ile Leu Thr Val
275 280 285
Val Leu Ser Trp Val Ser Phe Trp Ile Asn Lys Asp Ala Val Pro
290 295 300
Ala Arg Thr Ser Leu Gly Ile Thr Thr Val Leu Thr Met Thr Thr
305 310 315
Leu Ser Thr Ile Ala Arg Lys Ser Leu Pro Lys Val Ser Tyr Val
320 325 330
Thr Ala Met Asp Leu Phe Val Ser Val Cys Phe Ile Phe Val Phe
335 340 345
Ala Ala Leu Met Glu Tyr Gly Thr Leu His Tyr Phe Thr Ser Asn
350 355 360
Gln Lys Gly Lys Thr Ala Thr Lys Asp Arg Lys Leu Lys Asn Lys
365 370 375
Ala Ser Met Thr Pro Gly Leu His Pro Gly Ser Thr Leu Ile Pro
380 385 390

```
Met Asn Asn Ile Ser Val Pro Gln Glu Asp Asp Tyr Gly Tyr Gln
                395                 400                 405
Cys Leu Glu Gly Lys Asp Cys Ala Ser Phe Phe Cys Cys Phe Glu
                410                 415                 420
Asp Cys Arg Thr Gly Ser Trp Arg Glu Gly Arg Ile His Ile Arg
                425                 430                 435
Ile Ala Lys Ile Asp Ser Tyr Ser Arg Ile Phe Phe Pro Thr Ala
                440                 445                 450
Phe Ala Leu Phe Asn Leu Val Tyr Trp Val Gly Tyr Leu Tyr Leu
                455                 460                 465


<210> 23
<211> 235
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477105CD1

<400> 23
Met Gly Ser Val Gly Ser Gln Arg Leu Glu Glu Pro Ser Val Ala
  1               5                  10                  15
Gly Thr Pro Asp Pro Gly Val Val Met Ser Phe Thr Phe Asp Ser
                 20                  25                  30
His Gln Leu Glu Glu Ala Ala Glu Ala Ala Gln Gly Gln Gly Leu
                 35                  40                  45
Arg Ala Arg Gly Val Pro Ala Phe Thr Asp Thr Thr Leu Asp Glu
                 50                  55                  60
Pro Val Pro Asp Asp Arg Tyr His Ala Ile Tyr Phe Ala Met Leu
                 65                  70                  75
Leu Ala Gly Val Gly Phe Leu Leu Pro Tyr Asn Ser Phe Ile Thr
                 80                  85                  90
Asp Val Asp Tyr Leu His His Lys Tyr Pro Gly Thr Ser Ile Val
                 95                 100                 105
Phe Asp Met Ser Leu Thr Tyr Ile Leu Val Ala Leu Ala Ala Val
                110                 115                 120
Leu Leu Asn Asn Val Leu Val Glu Arg Leu Thr Leu His Thr Arg
                125                 130                 135
Ile Thr Ala Gly Tyr Leu Leu Ala Leu Gly Pro Leu Leu Phe Ile
                140                 145                 150
Ser Ile Cys Asp Val Trp Leu Gln Leu Phe Ser Arg Asp Gln Ala
                155                 160                 165
Tyr Ala Ile Asn Leu Ala Ala Val Gly Thr Val Ala Phe Gly Cys
                170                 175                 180
Thr Val Gln Gln Ser Ser Phe Tyr Gly His Arg Leu Ala Gln Pro
                185                 190                 195
Pro Pro Gly Thr Pro Pro His Glu Leu Trp Ser Pro Glu Arg Arg
                200                 205                 210
Gly Ala Ala Pro His Leu Val Thr Leu Arg Ala Ser Pro Ser Val
                215                 220                 225
Leu Ile Leu Arg Asp Cys Phe Ser Gln Thr
                230                 235

<210> 24
```

<211> 662
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482079CD1

<400> 24
Met Leu Lys Gln Ser Glu Arg Arg Arg Ser Trp Ser Tyr Arg Pro
1   5   10   15
Trp Asn Thr Thr Glu Asn Glu Gly Ser Gln His Arg Arg Ser Ile
                20   25   30
Cys Ser Leu Gly Ala Arg Ser Gly Ser Gln Ala Ser Ile His Gly
                35   40   45
Trp Thr Glu Gly Asn Tyr Asn Tyr Tyr Ile Glu Glu Asp Glu Asp
                50   55   60
Gly Glu Glu Glu Asp Gln Trp Lys Asp Asp Leu Ala Glu Glu Asp
                65   70   75
Gln Gln Ala Gly Glu Val Thr Thr Ala Lys Pro Glu Gly Pro Ser
                80   85   90
Asp Pro Pro Ala Leu Leu Ser Thr Leu Asn Val Asn Val Gly Gly
                95   100   105
His Ser Tyr Gln Leu Asp Tyr Cys Glu Leu Ala Gly Phe Pro Lys
                110   115   120
Thr Arg Leu Gly Arg Leu Ala Thr Ser Thr Ser Arg Ser Arg Gln
                125   130   135
Leu Ser Leu Cys Asp Asp Tyr Glu Glu Gln Thr Asp Glu Tyr Phe
                140   145   150
Phe Asp Arg Asp Pro Ala Val Phe Gln Leu Val Tyr Asn Phe Tyr
                155   160   165
Leu Ser Gly Val Leu Leu Val Leu Asp Gly Leu Cys Pro Arg Arg
                170   175   180
Phe Leu Glu Glu Leu Gly Tyr Trp Gly Val Arg Leu Lys Tyr Thr
                185   190   195
Pro Arg Cys Cys Arg Ile Cys Phe Glu Glu Arg Arg Asp Glu Leu
                200   205   210
Ser Glu Arg Leu Lys Ile Gln His Glu Leu Arg Ala Gln Ala Gln
                215   220   225
Val Glu Glu Ala Glu Glu Leu Phe Arg Asp Met Arg Phe Tyr Gly
                230   235   240
Pro Gln Arg Arg Arg Leu Trp Asn Leu Met Glu Lys Pro Phe Ser
                245   250   255
Ser Val Ala Ala Lys Ala Ile Gly Val Ala Ser Ser Thr Phe Val
                260   265   270
Leu Val Ser Val Val Ala Leu Ala Leu Asn Thr Val Glu Glu Met
                275   280   285
Gln Gln His Ser Gly Gln Gly Glu Gly Gly Pro Asp Leu Arg Pro
                290   295   300
Ile Leu Glu His Val Glu Met Leu Cys Met Gly Phe Phe Thr Leu
                305   310   315
Glu Tyr Leu Leu Arg Leu Ala Ser Thr Pro Asp Leu Arg Arg Phe
                320   325   330
Ala Arg Ser Ala Leu Asn Leu Val Asp Leu Val Ala Ile Leu Pro
                335   340   345
Leu Tyr Leu Gln Leu Leu Leu Glu Cys Phe Thr Gly Glu Gly His

```
                350                 355                 360
Gln Arg Gly Gln Thr Val Gly Ser Val Gly Lys Val Gly Gln Val
                365                 370                 375
Leu Arg Val Met Arg Leu Met Arg Ile Phe Arg Ile Leu Lys Leu
                380                 385                 390
Ala Arg His Ser Thr Gly Leu Arg Ala Phe Gly Phe Thr Leu Arg
                395                 400                 405
Gln Cys Tyr Gln Gln Val Gly Cys Leu Leu Leu Phe Ile Ala Met
                410                 415                 420
Gly Ile Phe Thr Phe Ser Ala Ala Val Tyr Ser Val Glu His Asp
                425                 430                 435
Val Pro Ser Thr Asn Phe Thr Thr Ile Pro His Ser Trp Trp Trp
                440                 445                 450
Ala Ala Val Ser Thr Phe Ala Leu Gly Phe Pro Ile Leu Phe Pro
                455                 460                 465
Ser Pro Val Ser Cys Ser Ser Leu Pro Trp Leu Ser Ala Thr Arg
                470                 475                 480
Leu Trp Leu Leu Ile Leu Val Phe Pro Pro Thr Pro Asn Arg Arg
                485                 490                 495
Ile Gln Leu Thr Lys Arg Arg Trp Met Ser Lys Val Val Glu Arg
                500                 505                 510
Glu Leu Ser Arg Ser Val Asn Ser Ser Ser His Met Ser Met Ala
                515                 520                 525
Val Ala Lys Asn Lys Arg Glu Asn Ala Ser Pro Ile Met Gln Thr
                530                 535                 540
Leu His Lys Phe Leu Phe Met Ala Phe Ala Gln Pro Ile Gly Gln
                545                 550                 555
Ser Lys Ser His Gly Gln Ala Ala Ser Gln Arg Ala Gly Gln Val
                560                 565                 570
Ser Ile Ser Thr Val Gly Tyr Gly Asp Met Tyr Pro Glu Thr His
                575                 580                 585
Leu Gly Arg Phe Phe Ala Phe Leu Cys Ile Ala Phe Gly Ile Ile
                590                 595                 600
Leu Asn Gly Met Pro Ile Ser Ile Leu Tyr Asn Lys Phe Ser Asp
                605                 610                 615
Tyr Tyr Ser Lys Leu Lys Ala Tyr Glu Tyr Thr Thr Ile Arg Arg
                620                 625                 630
Glu Arg Gly Glu Val Asn Phe Met Gln Arg Ala Arg Lys Lys Ile
                635                 640                 645
Ala Glu Cys Leu Leu Gly Ser Asn Pro Gln Leu Thr Pro Arg Gln
                650                 655                 660
Glu Asn


<210> 25
<211> 371
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55145506CD1

<400> 25
Met Asn Asp Glu Asp Tyr Ser Thr Ile Tyr Asp Thr Ile Gln Asn
  1               5                  10                  15
```

```
Glu Arg Thr Tyr Glu Val Pro Asp Gln Pro Glu Glu Asn Glu Ser
                20                  25                  30
Pro His Tyr Asp Asp Val His Glu Tyr Leu Arg Pro Glu Asn Asp
                35                  40                  45
Leu Tyr Ala Thr Gln Leu Asn Thr His Glu Tyr Asp Phe Val Ser
                50                  55                  60
Val Tyr Thr Ile Lys Gly Glu Glu Thr Ser Leu Ala Ser Val Gln
                65                  70                  75
Ser Glu Asp Arg Gly Tyr Leu Leu Pro Asp Glu Ile Tyr Ser Glu
                80                  85                  90
Leu Gln Glu Ala His Pro Gly Glu Pro Gln Glu Asp Arg Gly Ile
                95                  100                 105
Ser Met Glu Gly Leu Tyr Ser Ser Ala Gln Asp Gln Gln Leu Cys
                110                 115                 120
Ala Ala Glu Leu Gln Glu Asn Gly Ser Val Met Lys Glu Asp Leu
                125                 130                 135
Pro Ser Pro Ser Ser Phe Thr Ile Gln His Ser Lys Ala Phe Ser
                140                 145                 150
Thr Thr Lys Tyr Ser Cys Tyr Ser Asp Ala Glu Gly Leu Glu Glu
                155                 160                 165
Lys Glu Gly Ala His Met Asn Pro Glu Ile Tyr Leu Phe Val Lys
                170                 175                 180
Ala Gly Ile Asp Gly Glu Ser Ile Gly Asn Cys Pro Phe Ser Gln
                185                 190                 195
Arg Leu Phe Met Ile Leu Trp Leu Lys Gly Val Val Phe Asn Val
                200                 205                 210
Thr Thr Val Asp Leu Lys Arg Lys Pro Ala Asp Leu His Asn Leu
                215                 220                 225
Ala Pro Gly Thr His Pro Pro Phe Leu Thr Phe Asn Gly Asp Val
                230                 235                 240
Lys Thr Asp Val Asn Lys Ile Glu Glu Phe Leu Glu Glu Thr Leu
                245                 250                 255
Thr Pro Glu Lys Tyr Pro Lys Leu Ala Ala Lys His Arg Glu Ser
                260                 265                 270
Asn Thr Ala Gly Ile Asp Ile Phe Ser Lys Phe Ser Ala Tyr Ile
                275                 280                 285
Lys Asn Thr Lys Gln Gln Asn Asn Ala Ala Leu Glu Arg Gly Leu
                290                 295                 300
Thr Lys Ala Leu Lys Lys Leu Asp Asp Tyr Leu Asn Thr Pro Leu
                305                 310                 315
Pro Glu Glu Ile Asp Ala Asn Thr Cys Gly Glu Asp Lys Gly Ser
                320                 325                 330
Arg Arg Lys Phe Leu Asp Gly Asp Glu Leu Thr Leu Ala Asp Cys
                335                 340                 345
Asn Leu Leu Pro Lys Leu His Val Val Lys Thr His Leu Leu Thr
                350                 355                 360
Ser Ser Ser Asn Phe Leu Arg Asn Lys Tyr His
                365                 370
```

<210> 26
<211> 468
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 5950519CD1

<400> 26

```
Met Arg Gly Ser Pro Gly Asp Ala Glu Arg Arg Gln Arg Trp Gly
  1               5                  10                  15
Arg Leu Phe Glu Glu Leu Asp Ser Asn Lys Asp Gly Arg Val Asp
                 20                  25                  30
Val His Glu Leu Arg Gln Gly Leu Ala Arg Leu Gly Gly Gly Asn
                 35                  40                  45
Pro Asp Pro Gly Ala Gln Gln Gly Ile Ser Ser Glu Gly Asp Ala
                 50                  55                  60
Asp Pro Asp Gly Gly Leu Asp Leu Glu Glu Phe Ser Arg Tyr Leu
                 65                  70                  75
Gln Glu Arg Glu Gln Arg Leu Leu Leu Met Phe His Ser Leu Asp
                 80                  85                  90
Arg Asn Gln Asp Gly His Ile Asp Val Ser Glu Ile Gln Gln Ser
                 95                 100                 105
Phe Arg Ala Leu Gly Ile Ser Ile Ser Leu Glu Gln Ala Glu Lys
                110                 115                 120
Ile Leu His Ser Met Asp Arg Asp Gly Thr Met Thr Ile Asp Trp
                125                 130                 135
Gln Glu Trp Arg Asp His Phe Leu Leu His Ser Leu Glu Asn Val
                140                 145                 150
Glu Asp Val Leu Tyr Phe Trp Lys His Ser Thr Val Leu Asp Ile
                155                 160                 165
Gly Glu Cys Leu Thr Val Pro Asp Glu Phe Ser Lys Gln Glu Lys
                170                 175                 180
Leu Thr Gly Met Trp Trp Lys Gln Leu Val Ala Gly Ala Val Ala
                185                 190                 195
Gly Ala Val Ser Arg Thr Gly Thr Ala Pro Leu Asp Arg Leu Lys
                200                 205                 210
Val Phe Met Gln Val His Ala Ser Lys Thr Asn Arg Leu Asn Ile
                215                 220                 225
Leu Gly Gly Leu Arg Ser Met Val Leu Glu Gly Gly Ile Arg Ser
                230                 235                 240
Leu Trp Arg Gly Asn Gly Ile Asn Val Leu Lys Ile Ala Pro Glu
                245                 250                 255
Ser Ala Ile Lys Phe Met Ala Tyr Glu Gln Ile Lys Arg Ala Ile
                260                 265                 270
Leu Gly Gln Gln Glu Thr Leu His Val Gln Glu Arg Phe Val Ala
                275                 280                 285
Gly Ser Leu Ala Gly Ala Thr Ala Gln Thr Ile Ile Tyr Pro Met
                290                 295                 300
Glu Val Leu Lys Thr Arg Leu Thr Leu Arg Arg Thr Gly Gln Tyr
                305                 310                 315
Lys Gly Leu Leu Asp Cys Ala Arg Arg Ile Leu Glu Arg Glu Gly
                320                 325                 330
Pro Arg Ala Phe Tyr Arg Gly Tyr Leu Pro Asn Val Leu Gly Ile
                335                 340                 345
Ile Pro Tyr Ala Gly Ile Asp Leu Ala Val Tyr Glu Thr Leu Lys
                350                 355                 360
Asn Trp Trp Leu Gln Gln Tyr Ser His Asp Ser Ala Asp Pro Gly
                365                 370                 375
Ile Leu Val Leu Leu Ala Cys Gly Thr Ile Ser Ser Thr Cys Gly
                380                 385                 390
Gln Ile Ala Ser Tyr Pro Leu Ala Leu Val Arg Thr Arg Met Gln
```

```
                395                 400                 405
Ala Gln Ala Ser Ile Glu Gly Gly Pro Gln Leu Ser Met Leu Gly
                410                 415                 420
Leu Leu Arg His Ile Leu Ser Gln Glu Gly Met Arg Gly Leu Tyr
                425                 430                 435
Arg Gly Ile Ala Pro Asn Phe Met Lys Val Ile Pro Ala Val Ser
                440                 445                 450
Ile Ser Tyr Val Val Tyr Glu Asn Met Lys Gln Ala Leu Gly Val
                455                 460                 465
Thr Ser Arg
```

```
<210> 27
<211> 2229
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1687189CB1

<400> 27
gcctgagcgg ccgaactcgg cagctccaac ccaactcggc ttaactccgc ctcaccgagc 60
ccagtccaag actctgtgct ccctaggttt gcaacagctc tctgatcatc ttcttcaatt 120
cctgctagga tgccgtggca agcatttcgc agatttggtc aaaagctggt acgcagacgt 180
acactggagt caggcatggc tgagactcgc cttgccagat gcctaagcac cctggattta 240
gtggccctgg gtgtgggcag cacattgggt gcaggcgtgt atgtcctagc tggcgaggtg 300
gccaaagata aagcagggcc atccattgtg atctgctttt tggtggctgc cctgtcttct 360
gtgttggctg ggctgtgcta tgcggagttt ggtgcccggg ttccccgttc tggttcggca 420
tatctctaca gctatgtcac tgtgggtgaa ctctgggcct tcaccactgg ctggaacctc 480
atcctctcct atgtcattgg tacagccagt gtggcccggg cctggagctc tgcttttgac 540
aacctgattg ggaaccacat ctctaagact ctgcaggggt ccattgcact gcacgtgccc 600
catgtccttg cagaatatcc agatttcttt gctttgggcc tcgtgttgct gctcactgga 660
ttgttggctc tcggggctag tgagtcggcc ctggttacca aagtgttcac aggcgtgaac 720
cttttggttc ttgggttcgt catgatctct ggcttcgtta agggggacgt gcacaactgg 780
aagctcacag aagaggacta cgaattggcc atggctgaac tcaatgacac ctatagcttg 840
ggtcctctgg gctctggagg atttgtgcct ttcggcttcg agggaattct ccgtggagca 900
gcgacctgtt tctatgcatt tgttggtttc gactgtattg ctaccactgg agaagaagcc 960
cagaatcccc agcgttccat cccgatgggc attgtgatct cactgtctgt ctgctttttg 1020
gcgtattttg ctgtctcttc tgcactcacc ctgatgatgc cttactacca gcttcagcct 1080
gagagccctt tgcctgaggc atttctctac attggatggg ctcctgcccg ctatgttgtg 1140
gctgttggct ccctctgtgc tctttctacc agcctcctgg gctccatgtt ccccatgcct 1200
cgggtgatct acgcgatggc agaggatggc ctcctgttcc gtgtacttgc tcggatccac 1260
accggcacac gcaccccaat catagccacc gtggtctctg gcattattgc agcattcatg 1320
gcattcctct tcaaactcac tgatcttgtg gacctcatgt caattgggac cctgcttgct 1380
tactccctgg tgtcgatttg tgttctcatc ctcaggtatc aacctgatca ggagacaaag 1440
actggggaag aagtggagtt gcaggaggag gcaataacta ctgaatcaga gaagttgacc 1500
ctatggggac tatttttccc actcaactcc atccccactc cactctctgg ccaaattgtc 1560
tatgtttgtt cctcattgct tgctgtcctg ctgactgctc tttgcctggt gctggcccag 1620
tggtcagttc cattgctttc tggagacctg ctgtggactg cagtggttgt gctgctcctg 1680
ctgctcatta ttgggatcat tgtggtcatc tggagacagc cacagagctc cactcccctt 1740
cactttaagg tgcctgcttt gcctctcctc ccactaatga gcatctttgt gaatatttac 1800
cttatgatgc agatgacagc tggtacctgg gcccgatttg ggtctggat gctgattggc 1860
tttgctatct acttcggcta tgggatccag cacagcctgg aagagattaa gagtaaccaa 1920
ccctcacgca agtctagagc caaaactgta gaccttgatc ccggcactct ctatgtccac 1980
tcagtttgac atcgtcacac ctaaatgctg tctggtcccc tgcacaataa tggagagtac 2040
```

tcctgacccc agtgacagct agccctcccc tgtgatggtg gtggtggata ctaatacagt 2100
tctgtacgat gtgaaggatg tgtctttgct atttcttgtc tattttaacc cgtctgcttc 2160
taaatgatgt ctagctgctt accaacttta aaaaatgata ttaaaagaaa gtagaaaaat 2220
aaaaaaaaa 2229

<210> 28
<211> 7610
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7078207CB1

<400> 28
gcgccccgcc cccgcgcggg cgatgcccag cggcgcggcg ggctgcgggg cccggcgggg 60
cgcgcagagg agcgggccgc ggcgctgagg cggcggagcg tggccccgcc atgggcttcc 120
tgcaccagct gcagctgctg ctctggaaga acgtgacgct caaacgccgg agcccgtggg 180
tcctggcctt cgagatcttc atccccctgg tgctgttctt tatcctgctg ggctgcgac 240
agaagaagcc caccatctcc gtgaaggaag tctccttcta cacagcggcg cccctgacgt 300
ctgccggcat cctgcctgtc atgcaatcgc tgtgcccgga cggccagcga gacgagttcg 360
gcttcctgca gtacgccaac tccacggtca cgcagctgct tgagcgcctg gaccgcgtgg 420
tggaggaagg caacctgttt gacccagcgc ggcccagcct gggctcagag ctcgaggccc 480
tacgccagca tctggaggcc ctcagtgcgg gcccgggcac ctcggggagc cacctggaca 540
gatccacagt gtcttccttc tctctggact cggtggccag aaacccgcag gagctctggc 600
gtttcctgac gcaaaacttg tcgctgccca atagcacggc ccaagcactc ttggccgccc 660
gtgtggaccc gcccgaggtc taccacctgc tctttggtcc ctcatctgcc ctggattcac 720
agtctggcct ccacaagggt caggagccct ggagccgcct agggggcaat cccctgttcc 780
ggatggagga gctgctgctg gctcctgccc tcctggagca gctcacctgc acgccgggct 840
cgggggagct gggccggatc ctcactgtgc ctgagagtca gaagggagcc ctgcagggct 900
accgggatgc tgtctgcagt gggcaggctg ctgcgcgtgc caggcgcttc tctgggctgt 960
ctgctgagct ccggaaccag ctggacgtgg ccaaggtctc ccagcagctg ggcctggatg 1020
cccccaacgg ctcggactcc tcgccacagg cgccaccccc acggaggctg caggcgcttc 1080
tgggggacct gctggatgcc cagaaggttc tgcaggatgt ggatgtcctg tcggccctgg 1140
ccctgctact gccccagggt gcctgcactg gccggacccc cggaccccca gccagtggtg 1200
cgggtggggc ggccaatggc actggggcag gggcagtcat gggccccaac gccaccgctg 1260
aggagggcgc accctctgct gcagcactgg ccaccccgga cacgctgcag ggccagtgct 1320
cagccttcgt acagctctgg gccggcctgc agcccatctt gtgtggcaac aaccgcacca 1380
ttgaacccga ggcgctgcgg cggggcaaca tgagctccct gggcttcacg agcaaggagc 1440
agcggaacct gggcctcctc gtgcacctca tgaccagcaa ccccaaaatc ctgtacgcgc 1500
ctgcgggctc tgaggtcgac cgcgtcatcc tcaaggccaa cgagactttt gcttttgtgg 1560
gcaacgtgac tcactatgcc caggtctggc tcaacatctc ggcggagatc cgcagcttcc 1620
tggagcaggg caggctgcag caacacctgc gctggctgca gcagtatgta gcagagctgc 1680
ggctgcaccc cgaggcactg aacctgtcac tggatgagct gccgccggcc ctgagacagg 1740
acaacttctc gctgcccagt ggcatggccc tcctgcagca gctggatacc attgacaacg 1800
cggcctgcgg ctggatccag ttcatgtcca aggtgagcgt ggacatcttc aagggcttcc 1860
ccgacgagga gagcattgtc aactacaccc tcaaccaggc ctaccaggac aacgtcactg 1920
tttttgccag tgtgatcttc cagacccgga aggacggctc gctcccgcct cacgtgcact 1980
acaagatccg ccagaactcc agcttcaccg agaaaaccaa cgagatccgc cgcgcctact 2040
ggcggcctgg gcccaatact ggcggccgct tctacttcct ctacggcttc gtctggatcc 2100
aggacatgat ggagcgcgcc atcatcgaca cttttgtggg gcacgacgtg gtggagccag 2160
gcagctacgt gcagatgttc ccctacccct gctacacacg cgatgacttc ctgtttgtca 2220
ttgagcacat gatgccgctg tgcatggtga tctcctgggt ctactccgtg gccatgacca 2280
tccagcacat cgtggcggag aaggagcacc ggctcaagga ggtgatgaag accatgggcc 2340
tgaacaacgc ggtgcactgg gtggcctggt tcatcaccgg ctttgtgcag ctgtccatct 2400
ccgtgacagc actcaccgcc atcctgaagt acggccaggt gcttatgcac agccacgtgg 2460

tcatcatctg gctcttcctg gcagtctacg cggtggccac catcatgttc tgcttcctgg 2520
tgtctgtgct gtactccaag gccaagctgg cctcggcctg cggtggcatc atctacttcc 2580
tgagctacgt gccctacatg tacgtggcga tccgagagga ggtggcgcat gataagatca 2640
cggccttcga gaagtgcatc gcgtccctca tgtccacgac ggcctttggt ctgggctcta 2700
agtacttcgc gctgtatgag gtggccggcg tgggcatcca gtggcacacc ttcagccagt 2760
ccccggtgga gggggacgac ttcaacttgc tcctggctgt caccatgctg atggtggacg 2820
ccgtggtcta tggcatcctc acgtggtaca ttgaggctgt gcacccaggc atgtacgggc 2880
tgccccggcc ctggtacttc ccactgcaga agtcctactg gctgggcagt gggcggacag 2940
aagcctggga gtggagctgg ccgtgggcac gcaccccccg cctcagtgtc atggaggagg 3000
accaggcctg tgccatggag agccggcgct ttgaggagac ccgtggcatg gaggaggagc 3060
ccacccacct gcctctggtt gtctgcgtgg acaaactcac caaggtctac aaggacgaca 3120
agaagctggc cctgaacaag ctgagcctga acctctacga gaaccaggtg gtctccttct 3180
tgggccacaa cggggcgggc aagaccacca ccatgtccat cctgaccggc ctgttccctc 3240
caacgtcggg ttccgccacc atctacgggc acgacatccg cacggagatg gatgagatcc 3300
gcaagaacct gggcatgtgc ccgcagcaca atgtgctctt tgaccggctc acggtggagg 3360
aacacctctg gttctactca cggctcaaga gcatggctca ggaggagatc cgcagagaga 3420
tggacaagat gatcgaggac ctggagctct ccaacaaacg gcactcactg gtgcagacat 3480
tgtcgggtgg catgaagcgc aagctgtccg tggccatcgc cttcgtgggc ggctctcgcg 3540
ccatcatcct ggacgagccc acggcgggcg tggacccctа cgcgcgccgc gccatctggg 3600
acctcatcct gaagtacaag ccaggccgca ccatccttct gtccacccac cacatggatg 3660
aggctgacct gcttggggac cgcattgcca tcatctccca tgggaagctc aagtgctgcg 3720
gctccccgct cttcctcaag ggcacctatg gcgacgggta ccgcctcacg ctggtcaagc 3780
ggcccgccga gccgggggc ccccaagagc cagggctggc atccagcccc ccaggtcggg 3840
ccccgctgag cagctgctcc gagctccagg tgtcccagtt catccgcaag catgtggcct 3900
cctgcctgct ggtctcagac acaagcacgg agctctccta catcctgccc agcgaggccg 3960
ccaagaaggg ggctttcgag cgcctcttcc agcacctgga gcgcagcctg gatgcactgc 4020
acctcagcag cttcgggctg atggacacga ccctggagga agtgttcctc aaggtgtcgg 4080
aggaggatca gtcgctggag aacagtgagg ccgatgtgaa ggagtccagg aaggatgtgc 4140
tccctggggc ggagggcccg gcgtctgggg agggtcacgc tggcaatctg gcccggtgct 4200
cggagctgac ccagtcgcag gcatcgctgc agtcggcgtc atctgtgggc tctgcccgtg 4260
gcgacgaggg agctggctac accgacgtct atggcgacta ccgcccccctc tttgataacc 4320
cacaggaccc agacaatgtc agcctgcaag aggtggaggc agaggccctg tcgagggtcg 4380
gccagggcag ccgcaagctg gacggcgggt ggctgaaggt gcgccagttc cacgggctgc 4440
tggtcaaacg cttccactgc gcccgccgca actccaaggc actcttctcc cagatcttgc 4500
tgccagcctt cttcgtctgc gtggccatga ccgtggccct gtccgtcccg gagattggtg 4560
atctgccccc gctggtcctg tcaccttccc agtaccacaa ctacacccag ccccgtggca 4620
atttcatccc ctacgccaac gaggagcgcc gcgagtaccg gctgcggcta tcgcccgacg 4680
ccagccccca gcagctcgtg agcacgttcc ggctgccgtc gggggtgggt gccacctgcg 4740
tgctcaagtc tcccgccaac ggctcgctgg ggcccacgtt gaacctgagc agcggggagt 4800
cgcgcctgct ggcggctcgg ttcttcgaca gcatgtgtct ggagtccttc acacaggggc 4860
tgccactgtc caatttcgtg ccacccccac cctcgcccgc cccatctgac tcgccagcgt 4920
ccccggatga ggacctgcag gcctggaacg tctccctgcc gcccaccgct gggccagaaa 4980
tgtggacgtc ggcaccctcc ctgccgcgcc tggtacggga gcccgtccgc tgcacctgct 5040
ctgcgcaggg caccggcttc tcctgcccca gcagtgtggg cgggcacccg ccccagatgc 5100
gggtggtcac aggcgacatc ctgaccgaca tcaccggcca caatgtctct gagtacctgc 5160
tcttcacctc cgaccgcttc cgactgcacc ggtatggggc catcaccttt ggaaacgtcc 5220
tgaagtccat cccagcctca tttggcacca gggcccccacc catggtgcgg aagatcgcgg 5280
tgcgcagggc tgcccaggtt ttctacaaca acaagggcta tcacagcatg cccacctacc 5340
tcaacagcct caacaacgcc atcctgcgtg ccaacctgcc caagagcaag ggcaacccgg 5400
cggcttacgg catcaccgtc accaaccacc ccatgaataa gaccagcgcc agcctctccc 5460
tggattacct gctgcagggc acggatgtcg tcatcgccat cttcatcatc gtggccatgt 5520
ccttcgtgcc ggccagcttc gttgtcttcc tcgtggccga gaagtccacc aaggccaagc 5580
atctgcagtt tgtcagcggc tgcaacccca tcatctactg gctggcgaac tacgtgtggg 5640
acatgctcaa ctacctggtc cccgctacct gctgtgtcat catcctgttt gtgttcgacc 5700
tgccggccta cacgtcgccc accaacttcc ctgccgtcct ctccctcttc ctgctctatg 5760
ggtggtccat cacgcccatc atgtacccgg cctccttctg gttcgaggtc cccagctccg 5820

cctacgtgtt cctcattgtc atcaatctct tcatcggcat caccgccacc gtggccacct 5880
tcctgctaca gctcttcgag cacgacaagg acctgaaggt tgtcaacagt tacctgaaaa 5940
gctgcttcct cattttcccc aactacaacc tgggccacgg gctcatggag atggcctaca 6000
acgagtacat caacgagtac tacgccaaga ttggccagtt tgacaagatg aagtccccgt 6060
tcgagtggga cattgtcacc cgcggactgg tggccatggc ggttgagggc gtcgtgggct 6120
tcctcctgac catcatgtgc cagtacaact tcctgcggcg gccacagcgc atgcctgtgt 6180
ctaccaagcc tgtggaggat gatgtggacg tggccagtga gcggcagcga gtgctccggg 6240
gagacgccga caatgacatg gtcaagattg agaacctgac caaggtctac aagtcccgga 6300
agattggccg tatcctggcc gttgaccgcc tgtgcctggg tgtgcgtcct ggcgagtgct 6360
tcgggctcct gggcgtcaac ggtgcgggca agaccagcac cttcaagatg ctgaccggcg 6420
acgagagcac gacggggggc gaggccttcg tcaatggaca cagcgtgctg aaggagctgc 6480
tccaggtgca gcagagcctc ggctactgcc cgcagtgtga cgcgctgttc gacgagctca 6540
cggcccggga gcacctgcag ctgtacacgc ggctgcgtgg gatctcctgg aaggacgagg 6600
cccgggtggt gaagtgggct ctggagaagc tggagctgac caagtacgca gacaagccgg 6660
ctggcaccta cagcggcggc aacaagcgga agctctccac ggccatcgcc ctcattgggt 6720
acccagcctt catcttcctg gacgagccca ccacaggcat ggaccccaag gcccggcgct 6780
tcctctggaa cctcatcctc gacctcatca agacagggcg ttcagtggtg ctgacatcac 6840
acagcatgga ggagtgcgag gcgctgtgca cgcggctggc catcatggtg aacggtcgcc 6900
tgcggtgcct gggcagcatc cagcacctga agaaccggtt tggagatggc tacatgatca 6960
cggtgcggac caagagcagc cagagtgtga aggacgtggt gcggttcttc aaccgcaact 7020
tcccggaagc catgctcaag gagcggcacc acacaaaggt gcagtaccag ctcaagtcgg 7080
agcacatctc gctggcccag gtgttcagca agatggagca ggtgtctggc gtgctgggca 7140
tcgaggacta ctcggtcagc cagaccacac tggacaatgt gttcgtgaac tttgccaaga 7200
agcagagtga caacctggag cagcaggaga cggagccgcc atccgcactg cagtcccctc 7260
tcggctgctt gctcagcctg ctccggcccc ggtctgcccc cacggagctc cgggcacttg 7320
tggcagacga gcccgaggac ctggacacgg aggacgaggg cctcatcagc ttcgaggagg 7380
agcgggccca gctgtccttc aacacggaca cgctctgctg accacccaga gctgggccag 7440
ggaggacacg ctccactgac cacccagagc tggccaggg actcaacaat ggggacagaa 7500
gtcccccagt gcctgccagg gcctggagtg gaggttcagg accaaggggc ttctggtcct 7560
ccagcccctg tactcggcca tgtcctgcgg tcactgcggt tgccggccct 7610

<210> 29
<211> 2219
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1560619CB1

<400> 29
ggcagcatga gccgatcacc cctcaatccc agccaactcc gatcagtggg ctcccaggat 60
gccctggccc ccttgcctcc acctgctccc cagaatccct ccacccactc ttgggacccct 120
ttgtgtggat ctctgccttg gggcctcagc tgtcttctgg ctctgcagca tgtcttggtc 180
atggcttctc tgctctgtgt ctcccacctg ctcctgcttt gcagtctctc cccaggagga 240
ctctcttact cccctctca gctcctggcc tccagcttct tttcatgtgg tatgtctacc 300
atcctgcaaa cttggatggg cagcaggctg cctcttgtcc aggctccatc cttagagttc 360
cttatccctg ctctggtgct gaccagccag aagctacccc gggccatcca gacacctgga 420
aactcctccc tcatgctgca cctttgtagg ggacctagct gccatggcct ggggcactgg 480
aacacttctc tccaggaggt gtccggggca gtggtagtat ctgggctgct gcagggcatg 540
atggggctgc tggggagtcc cggccacgtg ttccccact gtgggcccct ggtgctggct 600
cccagcctgg ttgtggcagg gctctctgcc cacagggagg tagcccagtt ctgcttcaca 660
cactgggggt tggccttgct ggttatcctg ctcatggtgg tctgttctca gcacctgggc 720
tcctgccagt ttcatgtgtg cccctggagg cgagcttcaa cgtcatcaac tcacactcct 780
ctccctgtct tccggctcct ttcggtgctg atcccagtgg cctgtgtgtg gattgtttct 840
gcctttgtgg gattcagtgt tatccccag gaactgtctg cccccaccaa ggcaccatgg 900

atttggctgc ctcacccagg tgagtggaat tggcctttgc tgacgcccag agctctggct 960
gcaggcatct ccatggcctt ggcagcctcc accagttccc tgggctgcta tgccctgtgt 1020
ggccggctgc tgcatttgcc tcccccacct ccacatgcct gcagtcgagg gctgagcctg 1080
gaggggctgg gcagtgtgct ggccgggctg ctgggaagcc ccatgggcac tgcatccagc 1140
ttccccaacg tgggcaaagt gggtcttatc caggctggat ctcagcaagt ggctcactta 1200
gtggggctac tctgcgtggg gcttggactc tcccccaggt tggctcagct cctcaccacc 1260
atcccactgc ctgttgttgg tggggtgctg gggtgaccc aggctgtggt tttgtctgct 1320
ggattctcca gcttctacct ggctgacata gactctgggc gaaatatctt cattgtgggc 1380
ttctccatct tcatggcctt gctgctgcca agatggtttc gggaagcccc agtcctgttc 1440
agcacaggct ggagcccctt ggatgtatta ctgcactcac tgctgacaca gcccatcttc 1500
ctggctggac tctcaggctt cctactagag aacacgattc ctggcacaca gcttgagcga 1560
ggcctaggtc aagggctacc atctcctttc actgcccaag aggctcgaat gcctcagaag 1620
cccagggaga aggctgctca agtgtacaga cttcctttcc ccatccaaaa cctctgtccc 1680
tgcatccccc agcctctcca ctgcctctgc ccactgcctg aagaccctgg ggatgaggaa 1740
ggaggctcct ctgagccaga agagatggca gacttgctgc ctggctcagg ggagccatgc 1800
cctgaatcta gcagagaagg gtttaggtcc cagaaatgac cagaacgcct acttctgccc 1860
tggttaattt agccctaact ctcatctgct ggagagtcag ctcccaaact gttctttctt 1920
gtaggcagag gatatgtgtg tgtgtattac atgggactgt ctagaggttc catttcccaa 1980
tagggtgggt tgcctttcct tgtcttaatt aggcctaact gttccagagc agaggccatg 2040
atttagtgga ccatgaatga ttgagatttt gcctgtgtac tatcaatgcc acttgaacca 2100
cagcattcac tttaatactt actgagcatc tcccatgtgc aaggtcctgg aactacaggg 2160
ataagacagg gtccatgccg tctcaaggca tttacggttt aaaaagacct ttgtaatta 2219

<210> 30
<211> 1280
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2614283CB1

<400> 30
ccgcaggagc cgggccggag tgagcgcacc tcgcggggcc cctcggggca ggtgggtgag 60
cgccacccgg agtcccgcgc gcaactttca gggcgcactc ggcggggcgg ctgcgcggct 120
gccgggactc ggcgcgggac tgcatggagg ccaaggagaa gcagcatctg ttggacgcca 180
ggccggcaat ccggtcatac acgggatctc tgtggcagga aggggctggc tggattcctc 240
tgccccgacc tggcctggac ttgcaggcca ttgagctggc tgcccagagc aaccatcact 300
gccatgctca gaagggtcct gacagtcact gtgaccccaa gaaggggaag gcccagcgcc 360
agctgtatgt agcctctgcc atctgcctgt tgttcatgat cggagaagtc gttggtgggt 420
acctggcaca cagcttggct gtcatgactg acgcagcaca cctgctcact gactttgcca 480
gcatgctcat cagcctcttc tccctctgga tgtcctcccg gccagccacc aagaccatga 540
actttggctg gcagagagct gagatcttgg gagccctggt ctctgtactg tccatctggg 600
tcgtgacggg ggtactggtg tacctggctg tggagcggct gatctctggg gactatgaaa 660
ttgacggggg gaccatgctg atcacgtcgg gctgcgctgt ggctgtgaac atcataatgg 720
ggttgaccct tcaccagtct ggccatgggc acagccacgg caccaccaac cagcaggagg 780
agaaccccag cgtccgagct gccttcatcc atgtgatcgg cgactttatg cagagcatgg 840
gtgtcctagt ggcagcctat attttatact tcaagccaga atacaagtat gtagacccca 900
tctgcacctt cgtcttctcc atcctggtcc tggggacaac cttgaccatc ctgagagatg 960
tgatcctggt gttgatggaa gggaccccca agggcgttga cttcacagct gttcgtgatc 1020
tgctgctgtc ggtggagggg gtagaagccc tgcacagcct gcatatctgg gcactgacgg 1080
tggcccagcc tgttctgtct gtccacatcg ccattgctca gaatacagac gcccaggctg 1140
tgctgaagac agccagcagc cgcctccaag ggaagttcca cttccacacc gtgaccatcc 1200
agatcgagga ctactcggag gacatgaagg actgtcaggc atgccagggc ccctcagact 1260
gactgctcag ccaggcacca 1280

<210> 31
<211> 2727
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2667691CB1

<400> 31
cagtagtggt gggacggcac tagctgctgg ggcctgccgc cccgggagtg gctgcagcag 60
cgccaggaat cgaggatggt aaaatgaccc aggggaagaa gaagaaacgg gccgcgaacc 120
gcagtatcat gctggccaag aagatcatca ttaaggacgg aggcacgcct caaggaatag 180
gttctcctag tgtctatcat gcagttatcg tcatcttttt ggagtttttt gcttggggac 240
tattgacagc acccaccttg gtggtattac atgaaacctt tcctaaacat acatttctga 300
tgaacggctt aattcaagga gtaaagggtt tgttgtcatt ccttagtgcc ccgcttattg 360
gtgctctttc tgatgtttgg ggccgaaaat ccttcttgct gctaacggtg tttttcacat 420
gtgccccaat tcctttaatg aagatcagcc catggtggta ctttgctgtt atctctgttt 480
ctggggtttt tgcagtgact ttttctgtgg tatttgcata cgtagcagat ataacccaag 540
agcatgaaag aagtatggct tatggactgg tttcagcaac atttgctgca agtttagtca 600
ccagtcctgc aattggagct tatcttggac gagtatatgg ggacagcttg gtggtggtct 660
tagctacagc aatagctttg ctagatattt gttttatcct tgttgctgtg ccagagtcgt 720
tgcctgagaa aatgcggcca gcatcctggg gagcacccat ttcctgggaa caagctgacc 780
cttttgcgtc cttaaaaaaa gtcggccaag attccatagt gctgctgatc tgcattacag 840
tgtttctctc ctacctaccg gaggcaggcc aatattccag ctttttttta tacctcagac 900
agataatgaa attttcacca gaaagtgttg cagcgtttat agcagtcctt ggcattcttt 960
ccattattgc acagaccata gtcttgagtt tacttatgag gtcaattgga aataagaaca 1020
ccattttact gggtctagga tttcaaatat tacagttggc atggtatggc tttggttcag 1080
aaccttggat gatgtgggct gctggggcag tagcagccat gtctagcatc acctttcctg 1140
ctgtcagtgc acttgtttca cgaactgctg atgctgatca acagggtgtc gttcaaggaa 1200
tgataacagg aattcgagga ttatgcaatg gtctgggacc ggccctctat ggattcattt 1260
tctacatatt ccatgtggaa cttaaagaac tgccaataac aggaacagac ttgggaacaa 1320
acacaagccc tcagcaccac tttgaacaga attccatcat ccctggccct cccttcctat 1380
ttggagcctg ttcagtactg ctggctctgc ttgttgcctt gtttattccg gaacatacca 1440
atttaagctt aaggtccagc agttggagaa agcactgtgg cagtcacagc catcctcata 1500
atacacaagc gccaggagag gccaaagaac ctttactcca ggacacaaat gtgtgacgac 1560
tgaaatcagg aagatttttc tatcagcacc caggtcttag ttttcacctc tagttctgga 1620
tgtacattcc atttccatcc acagtgtact ttaagattgt cttaagaaat gtatctgcat 1680
gaactccgtg ggaactaaag gaagtgggaa cttagaacca gacagttttc caaagatgtt 1740
acaatttctt ttgaaaaacc ttttgtttat tagcaccaat ttcttgccac taagctattt 1800
gttttattat acatcccttta attaaaaact atatatgtaa cttcttagat attagcaaat 1860
gtctctgcta ccatttcctt aaggtgttga gctttaactc tatgctgact cagtgagaca 1920
cagtaggtag tatggttgtg gacctatttg ttttaacatt gtaaaatttt gagtcagatt 1980
ttaatattgt aaaatcttgg gtcaaataat tcaaagcctt aatgcagatg cactaaaaca 2040
aagaaatggt aaatgaattg tttgcattta aaaaaaaaaa ctcttaagaa aactgtacta 2100
aatctgaatc atgttttgag cttgtttgca gtacttttaa acattattca ctactgtttt 2160
tgaagtgaga aagtatcagc catttagcat ttaagttggg gtatttagag cctgtaatct 2220
aaatgctggc tcaaatttat tccccagcta cttcttatac cactattctt ttaatgtttg 2280
cataatcata agcacctcaa cacttgaata cataatctaa aaattatata gtaaagctgg 2340
tagccttgaa aatgtcagtg tgatatctat tatgtagata aatatatata gtggcctttc 2400
aggactgtca cagtaacact ttatttacag agctaatgtt tgtcctaaat tttcaggacc 2460
ctagaggaga gctttataca attaccgatg tgaatttctc taaagtgtat atttttgtgt 2520
ccagttatat tatttaaaaa agtgttactt tgtaaaaatt gtatataaag aactgtatag 2580
tttacactgt tttcatcttg tgtgtggtta ttgcttaatg ctttttaaac ttggaacact 2640
cactatggtt aaataaggtc ttaaaagaaa tgtaaatatt ctgttaataa agttaaatat 2700
tttaatgatt tttttttaa aaaaaaa 2727

<210> 32
<211> 1631
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3211415CB1

<400> 32
```
ttgcgcttga atgttcgttg actggcccgt cggttttaca aggcccggac aagcgctggg 60
gattccccgtt tgaggcgtca ctactgtcac tgccatcacc ccacggagcc acttctagag 120
gggagtagac ccggcccttc gccgggcaga gaagatgttg cccctgtcca tcaaagacga 180
tgaatacaaa ccacccaagt tcaatttgtt cggcaagatc tcgggctggt ttaggtctat 240
actgtccgac aagacttccc ggaacctgtt tttcttcctg tgcctgaacc tctctttcgc 300
ttttgtggaa ctactctacg gcatctggag caactgctta ggcttgattt ccgactcttt 360
tcacatgttt ttcgatagca ctgccatttt ggctggactg gcagcttctg ttatttcaaa 420
atggagagat aatgatgctt tctcctatgg gtatgttaga gcggaagttc tggctggctt 480
tgtcaatggc ctatttttga tcttcactgc tttttttatt ttctcagaag gagttgagag 540
agcattagcc cctccagatg tccaccatga gagactgctt cttgtttcca ttcttgggtt 600
tgtggtaaac ctaataggaa tatttgtttt caaacatgga ggtcatggac attctcatgg 660
ctctggtggc cacggacaca gtcattccct ctttaatggt gctctagatc aggcacatgg 720
ccatgtcgat cattgccata gccatgaagt gaaacatggt gctgcacata gccatgatca 780
tgctcatgga catggacact ttcattctca tgatggcccg tccttaaaag aaacaacagg 840
acccagcaga cagattttac aaggtgtatt tttacatatc ctagcagata cacttggaag 900
tattggtgta attgcttctg ccatcatgat gcaaaatttt ggtctgatga tagcagatcc 960
tatctgttca attcttatag ccattcttat agttgtaagt gttattcctc ttttaagaga 1020
atctgttgga atattaatgc agagaactcc tcccctatta gaaaatagtc tgcctcagtg 1080
ctatcagagg gtacagcagt tgcaaggagt ttacagttta caggaacagc acttctggac 1140
tttatgttct gacgtttatg ttgggacctt gaaattaata gtagcacctg atgctgatgc 1200
taggtggatt ttaagccaaa cacataatat ttttactcag gctggagtga gacagctcta 1260
cgtacagatt gactttgcag ccatgtagtg aatggaaaga aattatgcac cttttatgga 1320
ccaaattttt ctggcccaac ccgatgagat gaagcatttc aaacttgagg agaagagaga 1380
ctgcaggacg aggtggacag aaaaaccgtc agtaacaccg aggacatcta aaatctagtg 1440
atgaacacga tgacaaccaa agtagcacaa gagaagaggc aagtcacagg gaggggtttg 1500
ggaaccttat gtcacgctta agaactggga tgggaggttt taaacaaaaa aaaaaaaaaa 1560
aaaaggggcg gccgccgact attgaaccct tcgcccgggg aataaattcc ggcccggtac 1620
ctcgaggggg g                                                      1631
```

<210> 33
<211> 2673
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4739923CB1

<400> 33
```
ggacatttta aaagggccgg agattgcggg cgtcagtggc catggcggat acagcgacta 60
cagcatcggc ggcggcggct agtgccgcta gcgcctcgag cgatgcacct cctttccaac 120
tgggcaaacc ccgcttccag cagacgtcct tctatggccg cttcaggcac ttcttggata 180
tcatcgaccc tcgcacactc tttgtcactg agagacgtct cagagaggct gtgcagctgc 240
tggaggacta taagcatggg accctgcgcc cgggggtcac caatgaacag ctctggagtg 300
cacagaaaat caagcaggct attctacatc cggacaccaa tgagaagatc ttcatgccat 360
ttagaatgcc aggttatatt ccttttggga cgccaattgt agtcggtctt ctcttgccca 420
```

accagacact ggcatccact gtcttctggc agtggctgaa ccagagccac aatgcctgtg 480
tcaactatgc aaaccgcaat gcgaccaagc cttcacctgc atccaagttc atccagggat 540
acctgggagc tgtcatcagc gccgtctcca ttgctgtggg ccttaatgtc ctggttcaga 600
aagccaacaa gctcacccca gccacccgcc ttctcatcca gaggtttgtg ccgttccctg 660
ctgtagccag tgccaatatc tgcaatgtgg tcctgatgcg gtacggggag ctggaggaag 720
ggattgatgt cctggacagc gatggcaacc tcgtgggctc ctccaagatc gcagcccgac 780
acgccctgct ggagacggcg ctgacgcgag tggtcctgcc catgcccatc ctggtgctac 840
cccgatcgt catgtccatg ctggagaaga cggctctcct gcaggcacgc ccccggctgc 900
tcctccctgt gcaaagcctc gtgtgcctgg cagccttcgg cctggccctg ccgctggcca 960
tcagcctctt cccgcaaatg tcagagattg aaacatccca attagagccg gagatagccc 1020
aggccacgag cagccggaca gtggtgtaca acaaggggtt gtgagtgtgg tcagcggcct 1080
ggggacggag cactgtgcag ccggggagct gaggggcagg gccgtagact cacggctgca 1140
cctgcaggga gcagcacgcc aaccccagca gtcctgggcc ccctgggaga gtgctcaacc 1200
tacagtggag ggagactgac ccattcacat tttaacatag gcaagaggag ttctaacaca 1260
tttcgtacaa aaaaataatc aagtgcattt ctgggcctta tgtggggttg tcaaaactcc 1320
actcagcaca attatgtgtg aagctgaaaa attgtagagt gcccatgggg tagaagtaga 1380
atcctttat actttggttc ctttttatt tttatttttt atcagaatca aatctgagcc 1440
ttagtttcag ctgaccagaa gtggcaggag gacaggtgga ggcgagccag attaggcctg 1500
gagttgggct ggtttggtgg ccaggctggt aaatttagga tattacaatg gccagcccag 1560
atggctccct gggctggcat ggggagggga gagaaggtgg tctgcacccc acaggatgaa 1620
ctagccatga ctagggtcct caggcagtgg cccagggaat cagggagcac tggaggcctc 1680
tgcaagattc tgtgggcagc tggctctgaa tgtagccagc ccacatccat tccagagctg 1740
cagaaccagt ccctctgagt gaagtccagt gaccctggag ctagggtccc ccttcgtgga 1800
gctctaactg attcagggcc cctgaagtga ccccagctcc aggcaggaaa ccccgagaag 1860
gaatggtgct tggcaggaac catggacctg cacttggcct cttcgggaag atcctccttc 1920
caggccccag gctggatgct gggttctggg gctgagagtg gggctagact gggctggctg 1980
ccttctgcgg agcttctcca gccaccaagg ctgcctgccc catcccatcc ctttctaagc 2040
aggaaggtct catgcctgag aatgcctagg ccagctcctt agacacctat cagagaagca 2100
gcatcaacct aggagcagtg ggccctggct ctgtcactaa atggcagtga gatgtcagac 2160
aaattcattc ccttctctca gcctccactc cctgtctgaa accagaagac tggatcaagg 2220
ggttcccact ggctcctcca gcaagacctc gtctttgctt gtcctgctca gatgctggtc 2280
atcctgggca tgtccccagt gtggactctg gactgggaag gggcaggcc cctttggacc 2340
tgcagttggc ctcagcagaa ggccttgcct tgtgtatgtg actccatatc ccgggagcag 2400
ttgacctttg ccaaacactt tacagttctg gaggaggagg taacatagat gcctgggcct 2460
gatggtgggg ccatacccat gtgtcgcctc tcactctggc agcctcagag gccccttgct 2520
gctggctccc atctccctcc catttgcaga ccaggaagga agagcaagct gtacaaaggg 2580
aagcagagcc tggggtgggt gtgagcaggg tgacccctca tctgaaaggc ccaaaccagg 2640
gggaagcacc agcctcagtg cagcccctc ctg 2673

<210> 34
<211> 3958
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55030459CB1

<400> 34
atggcccgcc agccggagga agaggagacg gccgtggccc gggcgcggcg gccgcccctc 60
tggctgctct gcctggtcgc gtgctggctc ctgggcgccg gggccgaagc cgacttctcc 120
atcctggacg aggcgcaagt gctggcgagc cagatgcgga ggctggcggc cgaggagctg 180
ggggtcgtca ccatgcagcg gatattcaac tcctttgttt acactgagaa aatctcaaat 240
ggagaaagtg aagtacagca gctagccaaa aaaatccgag agaagttcaa ccgttacttg 300
gatgtggtca atcggaacaa gcaagttgta gaagcatcct atacggctca cctaacctct 360
cccctaactg caattcaaga ctgctgtact atcccacctt ccatgatgga attcgatggg 420

aactttaata ccaatgtgtc tagaacaatt agttgtgatc gactttctac tactgttaat 480
agccgggcct tcaatccagg acgagactta aattcagttc ttgcagacaa cctgaaatcc 540
aaccctggaa ttaagtggca atatttcagt tcagaagaag gaattttcac tgttttccca 600
gcacacaagt tccggtgtaa gggcagctac gaacaccgca gtagacccat ctacgtctct 660
acagtccggc cgcagtcaaa gcacatagta gtgattctgg accacggggc ttcagtcaca 720
gacactcagc ttcagattgc caaggacgct gctcaggtca tcctcagcgc catcgatgaa 780
catgacaaga tttctgtgtt aactgtggca gataccgtcc ggacttgctc actagaccag 840
tgctataaga ccttcttgtc tccagccacc agtgagacaa aaaggaaaat gtccaccttt 900
gttagcagcg tgaagtcttc agacagtcct acccagcacg cagtgggatt ccaaaaggca 960
tttcagctga ttcgaagtac aaacaataac acaaagttcc aagcaaatac agacatggtc 1020
atcatttacc tgtcagctgg cattacatca aaggactctt cggaagaaga taaaaaagcg 1080
actctccaag tcatcaatga agaaaatagc tttctaaaca actctgtaat gattctcacc 1140
tatgccctca tgaacgatgg ggtgactggt ttgaaagagc tggcttttct gagggatcta 1200
gctgaacaga attcagggaa gtacggtgtg ccagaccgga cggccttgcc tgtgattaag 1260
ggcagcatga tggtgctgaa tcagttgagc aacctggaga ccacagtggg caggttctac 1320
acaaaccttc ccaaccggat gattgatgaa gccgtcttca gcctgccctt ctctgatgag 1380
atgggagatg gtttgataat gactgtgagt aaaccctgtt attttggaaa cctacttctg 1440
ggaattgtag gtgtggacgt gaatctggct tacattcttg aagacgtgac gtattaccaa 1500
gactctttgg cttcctatac ttttctcata gacgacaaag gatatacact tatgcaccca 1560
tctcttacca ggccatattt attgtcagag cccccacttc atactgacat catacattat 1620
gaaaatattc caaaatttga attagttcgg caaaatatcc taagcctccc tctgggcagc 1680
cagattatcg cagtccctgt gaactcatcc ctgtcttggc acataaacaa gctgagagaa 1740
actggaaagg aagcctacaa tgttagctat gcctggaaga tggtacaaga cacttccttt 1800
attctgtgta ttgtggtgat acaaccagaa atacctgtga aacaactgaa gaacctcaac 1860
actgttccca gcagcaagct gctgtaccac cggctggatc tccttggcca gcccagtgct 1920
tgcctccact tcaaacagct ggcaacccta gaaagtccca ccatcatgct gtctgctggc 1980
agcttttcct ccccctatga gcacctcagc cagccagaga caaagcgcat ggtagagcac 2040
tacaccgcct atctcagcga caacacccgc ctcattgcta acccgggcct caaattctct 2100
gtcagaaatg aagtaatggc taccagccac gtcacagatg aatggatgac acaaatggaa 2160
atgagtagcc tgaacactta cattgtccgc cgttacatag caacacccaa tggcgtcctc 2220
agaatttatc ctggttccct catggacaaa gcatttgatc ccactaggag acaatggtat 2280
ctccatgcag tagctaatcc agggttgatt tctttgactg gtccttactt agatgttgga 2340
ggagctggtt atgttgtgac aatcagtcac acaattcatt catccagtac acagctgtct 2400
tctgggcaca ctgtggctgt gatgggcatt gacttcacac tcagatactt ctacaaagtt 2460
ctgatggacc tattacctgt ctgtaaccaa gatggtggca acaaaataag gtgcttcata 2520
atggaggaca ggggttatct ggtggcgcac ccgactctca tcgaccccaa aggacatgca 2580
cctgtggagc agcagcacat cacccacaag gagcccctgg tagcaaatga tatcctcaac 2640
cacccccaact ttgtaaagaa aaacctgtgc aacagcttca gtgacagaac ggtccagagg 2700
ttttataaat tcaacaccag ccttgcgggg gatttgacga accttgtgca tggcagccac 2760
tgttccaaat acagattagc aaggatccca ggaaccaacg cgtttgttgg cattgtcaac 2820
gaaacctgcg actctcttgc cttctgtgcc tgcagcatgg tggaccgact ctgtctcaac 2880
tgtcaccgaa tggaacaaaa tgaatgtgaa tgtccttgtg agtgccctct agaggtcaat 2940
gagtgcactg gcaacctcac caatgcagag aaccgaaacc ccagctgcga ggtccaccag 3000
gagccggtga catacacagc tattgaccct ggcctgcaag atgctcttca ccagtgtgtc 3060
aacagcaggt gcagtcagag gctggaaagt ggggactgtt ttggggtgct ggattgtgaa 3120
tggtgcatgg tggacagtga tggaaagact cacctggaca aaccctactg tgccccccag 3180
aaagaatgct tcgggggggat tgtgggagcc aaaagtccct acgttgatga catgggagca 3240
ataggtgatg aggtgatcac attaaacatg attaaaagcg cccctgtggg tcctgtggct 3300
ggagggatca tgggatgcat catggtcttg gtcctggcgg tgtatgccta ccgccaccag 3360
attcatcgcc ggagccatca gcatatgtct cctcttgctg cccaagaaat gtcagtgcgt 3420
atgtccaacc tggagaatga cagagatgaa agggacgacg acagccacga agacagaggc 3480
atcatcagca acactcggtt tatagctgcg gtcatcgaac gacatgcaca cagtccagaa 3540
agaaggcgcc gctactgggg tcgatcagga acagaaagtg atcatggtta cagcaccatg 3600
agcccacagg aggacagtga aaatcctcca tgcaacaatg acccccttgtc agccggggtc 3660
gatgtgggaa accatgatga ggacttagac ctggataccc cccctcagac tgctgcccta 3720
ctaagtcaca agttccacca ctaccggtca caccacccta cacttcatca tagccaccac 3780

ttacaggcgg ccgtcacggt acacactgtc gatgcagaat gctaacaatc tcctcacctc 3840
cacgccaaga tgagatctgg gagctacaga atgttctgga aagaaaaaga accggcttaa 3900
aacccacaag cagagacctc ccttgtgttt gtgctttgtg cagagttgtt tgagtcat 3958

<210> 35
<211> 2000
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6113039CB1

<400> 35
gctcaggaca atgaaattct tcagttacat tctggtttat cgccgatttc tcttcgtggt 60
tttcactgtg ttggttttac tacctctgcc catcgtcctc cacaccaagg aagcagaatg 120
tgcctacaca ctctttgtgg tcgccacatt ttggctcaca gaagcattgc ctctgtcggt 180
aacagctttg ctacctagtt taatgttacc catgtttggg atcatgcctt ctaagaaggt 240
ggcatctgct tatttcaagg attttcactt actgctaatt ggagttatct gtttagcaac 300
atccatagaa aaatggaatt tgcacaagag aattgctctg aaaatggtga tgatggttgg 360
tgtaaatcct gcatggctga cgctggggtt catgagcagc actgcctttt tgtctatgtg 420
gctcagcaac acctcgacgg ctgccatggt gatgcccatt gcggaggctg tagtgcagca 480
gatcatcaat gcagaagcag aggtcgaggc cactcagatg acttacttca acggatcaac 540
caaccacgga ctagaaattg atgaaagtgt taatggacat gaaataaatg agaggaaaga 600
gaaaacaaaa ccagttccag gatacaataa tgatacaggg aaaatttcaa gcaaggtgga 660
gttggaaaag aactcaggca tgagaaccaa atatcgaaca aagaagggcc acgtgacacg 720
taaacttacg tgtttgtgca ttgcctactc ttctaccatt ggtggactga caacaatcac 780
tggtacctcc accaacttga tctttgcaga gtatttcaat acacgctatc ctgactgtcg 840
ttgcctcaac tttggatcat ggtttacgtt ttccttccca gctgccctta tcattctact 900
cttatcctgg atctggcttc agtggctttt cctaggattc aattttaagg agatgttcaa 960
atgtggcaaa accaaaacag tccaacaaaa agcttgtgct gaggtgatta agcaagaata 1020
ccaaaagctt gggccaataa ggtatcaaga aattgtgacc ttggtcctct tcattataat 1080
ggctctgcta tggtttagtc gagaccccgg atttgttcct ggttggtctg cactttttc 1140
agagtaccct ggttttgcta cagattcaac tgttgcttta cttataggc tgctattctt 1200
tcttatccca gctaagacac tgactaaaac tacacctaca ggagaaattg ttgcttttga 1260
ttactctcca ctgattactt ggaaagaatt ccagtcattc atgccctggg atatagccat 1320
tcttgttggt ggagggtttg ccctggcaga tggttgtgag gagtctggat tatctaagtg 1380
gataggaaat aaattatctc ctctgggttc attaccagca tggctaataa ttctgatatc 1440
ttctttgatg gtgacatctt taactgaggt agccagcaat ccagctacca ttacactctt 1500
tctcccaata ttatctccat tggccgaagc cattcatgtg aaccctcttt atattctgat 1560
accttctact ctgtgtactt catttgcatt cctcctacca gtagcaaatc cacccaatgc 1620
tattgtcttt tcatatggtc atctgaaagt cattgacatg gttaaagctg gacttggtgt 1680
caacattgtt ggtgttgctg tggttatgct tggcatatgt acttggattg tacccatgtt 1740
tgacctctac acttaccctt cgtgggctcc tgctatgagt aatgagacca tgccataata 1800
agcacaaaat ttctgactat cttgcggtaa tttctggaag acattaatga ttgactgtaa 1860
aatgtggctc taaataacta atgacacaca tttaaatcag ttatggtgta gctgctgcaa 1920
ttcccgtgaa tacccgaaac ctgctgttat aactcagagt ccatatttgt tattgcagtg 1980
caactaaaga gcatctatgt 2000

<210> 36
<211> 1997
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7101781CB1

<400> 36
ctgctcgcca ctggccggcg cgctcccggc gcacggagca cactcgcgct cccggcgcac 60
ggagcacact cgcgctccgg gactgaaacc tgagcagccg tagcagccga atttgggagc 120
atatccttgt cactgcagcc agaaagccct tcgatcccca tcagagaggt cacatgagcc 180
ccgaggtcac ctgcccgcgg aggggccacc tgcctcgctt ccacccgagg acctgggttg 240
agcccgtggt ggcatcgtcc caggtggctg cctccctcta cgatgcgggg ctactcctcg 300
tggtgaaggc gtcctacgga accggaggct cctccaacca cagtgccagc ccatcgcccc 360
ggggggctct agaggaccaa cagcagagag ccatctccaa tttctacatt atctacaacc 420
ttgtggtggg cctgtccccc ctgctgtccg cctacgggct gggatggctc agcgaccgct 480
accaccgaaa gatctccatc tgcatgtcgc tgctgggctt cctgctctcc cgcctcgggc 540
tgctgctcaa ggtgctgctg gactggccag tggaggtgct gtacggggcg gcggcgctga 600
acgggctatt cggcggcttc tccgccttct ggtccggggt catggcgctg ggatcgctgg 660
gctcctccga gggccgccgc tctgtgcgcc tcatcctcat tgacctgatg ctgggcttgg 720
cggggttctg cgggagcatg gcttccgggc atctcttcaa gcagatggct gggcactctg 780
ggcagggcct gatactgacg gcctgcagcg tgagctgtgc ctcgtttgcc ctgctctaca 840
gccttttggt gctaaaggtc cctgagtcgg tggccaaacc cagccaggag ctccccgccg 900
tggataccgt gtctggcacg gttggcacat accgcactct ggatcctgat cagttggacc 960
aacagtatgc agtggggcac cctccatctc ctggaaaagc aaaaccccat aaaaccacca 1020
ttgccttgct ctttgtgggt gctatcatat atgacctggc ggtggtgggc acagtggacg 1080
tgatccctct ttttgtgctg agggagcctc tcggttggaa ccaagtgcag gtgggctatg 1140
gtatggctgc agggtacacc atcttcatca ccagcttcct gggtgtcctg gtcttctccc 1200
gctgctttcg ggacaccacc atgatcatga ttgggatggt ctcctttggg tcaggagccc 1260
tcctcttggc ttttgtgaaa gagacataca tgttctatat tgctcgagcc gtcatgctgt 1320
ttgctctcat ccccgtcaca accatccgat cagctatgtc caaactcata aagggctcct 1380
cttatggaaa ggtgttcgtc atactgcagc tgtccttggc tctgaccggc gtggtgacat 1440
ccaccttgta caacaagatc taccagctca ccatggacat gtttgtgggc tcctgctttg 1500
ctctctcctc ctttctctcc ttcctggcca tcattccaat tagcatcgtg gcctataaac 1560
aagtcccatt gtcaccatat ggagacatca tagagaaatg aagatgctta cctgcaggaa 1620
ctgaaaacat cagccatggc caggccccca gaagacaaaa gaagggacca gggaactggt 1680
gacctaagca acccactgct taagaaacct gcgttccagc cagagttggc ctcagaatga 1740
cctgctctgg ctcagggatc cctggtggat ggggaaaagc actttcctgg tgatggaaaa 1800
acgttctcag ctttaagaca cccccattag gcagacactg ggttctgtga cagcagagca 1860
tgaccttaag ggttacaggg aggctgcaca cagcagtccc aggccctgtg aggggcttca 1920
gactccagct acagcgagcc tgcccctttt cttcaaggga ctgtcttgag ggctccaaag 1980
tatagctaac tagtcac 1997

<210> 37
<211> 3069
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473036CB1

<400> 37
atgcaaccag ccagagggcc cctggcttca gaacctagga ctgtactggt tctgagattc 60
tgtgcaagcc tcatggaaat gaagctgcca ggccaggaag ggtttgaagc ctccagtgct 120
cctagaaata ttccttcagg ggagctggac agcaaccctg accctggcac cggccccagc 180
cctgatggcc cctcagacac agagagcaag gaactgggag tacccaaaga ccctctgctc 240
ttcattcagc tgaatgagct gctgggctgg ccccaggcgc tggagtggag agagacaggc 300
aggtgggtac tgtttgagga gaagttggag gtggctgcag gccggtggag tgccccccac 360
gtgcccaccc tggcactgcc cagcctccag aagctccgca gcctgctggc cgagggcctt 420
gtactgctgg actgcccagc tcagagcctc ctggagctcg tgggctctac tcatccaaga 480

aaggcttctg acaatgagga agccccccctg agggaacagt gtcagaaccc cctgagacag 540
aagctacctc caggagctga ggcagggact gtgctggcag gggagctggg cttcctggca 600
cagccactgg gagcctttgt tcgactgcgg aaccctgtgg tactggggtc ccttactgag 660
gtgtccctcc caagcaggtt tttctgcctt ctcctgggcc cctgtatgct gggaaagggc 720
taccatgaga tgggacgggc agcagctgtc ctcctcagtg acccgcaatt ccagtggtca 780
gttcgtcggg ccagcaacct tcatgacctt ctggcagccc tggatgcatt cctagaggag 840
gtgacagtgc ttccccccagg tcggtgggac ccaacagccc ggattccccc gcccaaatgt 900
ctgccatctc agcacaaaag gcttccctcg caacagcggg agatcagagg tcccgccgtc 960
ccgcgcctga cctcggctga ggacaggcac cgccatgggc cacacgcaca cagcccggag 1020
ttgcagcgga ccggcagcga tttcttggac gccctgcatc tccagtgctt ctcggccgta 1080
ctctacattt acctggccac tgtcactaat gccatcactt ttgggggtct gctgggagat 1140
gccactgatg gtgcccaggg agtgctggaa agtttcctgg gcacagcagt ggctggagct 1200
gccttctgcc tgatggcagg ccagcccctc accattctga gcagcacggg gccagtgctg 1260
gtctttgagc gcctgctctt ctctttcagc agagattaca gcctggacta cctgcccttc 1320
cgcctatggg tgggcatctg ggtggctacc ttttgcctgg tgctggtggc cacagaggcc 1380
agtgtgctgg tgcgctactt cacccgcttc actgaggaag gtttctgtgc cctcatcagc 1440
ctcatcttca tctacgatgc tgtgggcaaa atgctgaact tgacccatac ctatcctatc 1500
cagaagcctg ggtcctctgc ctacgggtgc ctctgccaat acccaggccc aggaggaaat 1560
gagtctcaat ggataaggac aaggccaaaa gacagagacg acattgtaag catggactta 1620
ggcctgatca atgcatcctt gctgccgcca cctgagtgca cccggcaggg aggccaccct 1680
cgtggccctg gctgtcatac agtcccagac attgccttct tctcccttct cctcttcctt 1740
acttctttct tctttgctat ggccctcaag tgtgtaaaga ccagccgctt cttcccctct 1800
gtggtgcgca aagggctcag cgacttctcc tcagtcctgg ccatcctgct cggctgtggc 1860
cttgatgctt tcctgggcct agccacacca aagctcatgg tacccagaga gttcaagccc 1920
acactccctg ggcgtggctg gctggtgtca ccttttggag ccaacccctg gtggtggagt 1980
gtggcagctg ccctgcctgc cctgctgctg tctatcctca tcttcatgga ccaacagatc 2040
acagcagtca tcctcaaccg catggaatac agactgcaga agggagctgg cttccacctg 2100
gacctcttct gtgtggctgt gctgatgcta ctcacatcag cgcttggact gccttggtat 2160
gtctcagcca ctgtcatctc cctggctcac atggacagtc ttcggagaga gagcagagcc 2220
tgtgcccccg ggagcgcccc caacttcctg ggtatcaggg aacagaggct gacaggcctg 2280
gtggtgttca tccttacagg agcctccatc ttcctggcac ctgtgctcaa gttcattcca 2340
atgcctgtgc tctatggcat cttcctgtat atgggggtgg cagcgctcag cagcattcag 2400
ttcactaata gggtgaagct gttgttgatg ccagcaaaac accagccaga cctgctactc 2460
ttgcggcatg tgcctctgac cagggtccac ctcttcacag ccatccagct tgcctgtctg 2520
gggctgcttt ggataatcaa gtctacccct gcagccatca tcttcccccct catgttgctg 2580
ggccttgtgg gggtccgaaa ggccctggag agggtcttct caccacagga actcctctgg 2640
ctggatgagc tgatgccaga ggaggagaga agcatccctg agaaggggct ggagccagaa 2700
cactcattca gtggaagtga cagtgaagat tcagagctga tgtatcagcc aaaggctcca 2760
gaaatcaaca tttctgtgaa ttagctggag taggagtctg ggagtggaga ccccaggaaa 2820
cagcatgagt tcacaggtgc ttactcagga agtcaggaca tttttggcct ttggcttaac 2880
ttccagatgc tcagtcggct tggggaagga ctgaagggca gctgccaaga cctcagttac 2940
ctcctgacct gagggtggag agtggcagga agcaagcatg tttgctgtgc acttaggaaa 3000
ggctggtgag ccagagggac tgatcaggcc ccattcactc tctactcatt aaaaggtcct 3060
gagccacaa 3069

<210> 38
<211> 2241
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476943CB1

<400> 38
gccggcggcc cccgctcccg gatccccagc gccctggcca agaagcttcc tcggctcccc 60

ctcttccctc tccctgacac ggttgtgcag agggcgcggt ggctcaggcc ctggcaacca 120
ccattctact ttttgtgtct atgagtttga ctaccctaag gacctcacat ggcgagtaac 180
ccatgggcca ggtagcgttc tatgccaacc ttgaatgcca tcaggaagtc actggacagc 240
aaactcttcc aagatcataa cttggctgtt ggagcaacct ggaaaagaag aaaaaagaaa 300
aaccatggca aaagtaaata gagctcggtc tacctcccct ccagatggag gctggggctg 360
gatgattgtg gctggctgtt tccttgttac catctgcaca cgggcagtca caagatgtat 420
ctcaattttt tttgtggagt tccagacata cttcactcag gattacgcac aaacggcatg 480
gatccattcc attgtagatt gtgtgaccat gctctgtgct ccacttggga gtgttgtcag 540
taaccattta tcctgtcaag tgggaatcat gctgggtggc ttgcttgcat ctactggact 600
catcctgagc tcatttgcca cgagtctgaa gcatctctac ctcactctgg gagttcttac 660
aggtcttgga tttgcacttt gttactctcc agctattgcc atggttggca agtacttcag 720
cagacggaaa gcccttgctt atggtatcgc catgtcagga agtggcattg gcaccttcat 780
cctggctcct gtggttcagc tccttattga acagttttcc tggcggggag ccttactcat 840
tcttgggggc tttgtcttga atctctgtgt atgtggtgcc ttgatgaggc caattactct 900
taaagaggac cacacaactc cagagcagaa ccatgtgtgt agaactcaga aagaagacat 960
taagcgggtg tctccctatt catctttgac caaagaatgg gcacagactt gcctctgttg 1020
ctgtttgcag caagagtaca gtttttttact catgtcagac tttgttgtgt tagccgtctc 1080
cgttctgttt atggcttatg gctgcagccc tctctttgtg tacttggtgc cttatgcttt 1140
gagtgttgga gtgagtcatc agcaagctgc ttttcttatg tccatacttg gagtgattga 1200
cattattggc aatatcacat ttggatggct gaccgacaga aggtgtctga agaattacca 1260
gtatgtttgc tacctctttg ccgtgggaat ggatgggctc tgctatctct gcctcccaat 1320
gcttcaaagt ctccctctgc tcgtgccttt ctcttgtacc tttggctact ttgatggtgc 1380
ctatgtgact ttgatcccag tagtgaccac agagatagtg gggaccacct ctttgtcatc 1440
agcgcttggt gtggtatact tccttcacgc agtgccatac ttggtgagcc cacccatcgc 1500
aggacggctg gtagatacca ccggcagcta cactgcagca ttcctcctct gtggattttc 1560
aatgatattt agttctgtgt tgcttggctt tgctagactt ataaagagaa tgagaaaaac 1620
ccagttgcag ttcattgcca aagaatctga tcctaagctg cagctatgga ccaatggatc 1680
agtggcttat tctgtggcaa gagaattaga tcagaaacat ggggagcctg tggctacagc 1740
agtgcctggc tacagcctca catgaccaaa ggccttgagc cccagaatct tcaggtttga 1800
gagaggtggg gccaccagat tcttcatgtt tctgaaactt tttattttgg cagaaggatt 1860
gccttccaag gaaattatta ttattgtttt gttaacatat taatatttat aagggaaaac 1920
agcacataat aaggaaagct ggactagccc agagccttct catttgggat ttgtgctcat 1980
aactgaactc gtatctttgg tcaatgggca tagctctgta agaaatgtaa ggacacagct 2040
gatataatta gctgtaatta gggataattt caaagcataa ccaaagcaga tgacactggg 2100
cagcagcttt gttccagtct caggcccttc atgttccctc ctcagaaaga aatggaacat 2160
taacgtggta gctttggtta cttggtctgg ttagagaagg aggccagtga gtggggggtg 2220
aagtgaaaag caaataaagt a 2241

<210> 39
<211> 1593
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8003355CB1

<400> 39
ccttggagct gttgtcccac ccctgtcact gcagagagct gaggcaccat gcatgggggc 60
caggggccgc tgctcctcct gctgctgctg gctgtctgcc tgggggccca gggccggaac 120
caggaggagc gcctgctcgc agacctgatg caaaactacg accccaacct gcggcccgcg 180
gaacgagact cggatgtggt caatgtcagc ctgaagctaa ccctcaccaa cctcatctcc 240
ctgaacgagc gagaggaagc cctcaccacc aatgtctgga tagaggtgca gtggtgcgac 300
tatcgcctgc gccgggatcc gcgagactac gaaggcctgt gggtgctgag ggtgccgtcc 360
accatggtgt ggcggccgga tatcgtgctg gagaacaacg cggacggtgt cttcgaggtg 420
gccctctact gcaatgtgct cgtgtcccct gacggctgta tctactggct gccgcctgcc 480

atcttccgtt ccgcctgctc tatctcagtc acctacttcc ccttcgactg gcagaactgc 540
tcccttatct tccagtccca gacttacagc accaatgaga ttgatctgca gctgagtcag 600
gaagatggcc agaccatcga gtggattttc attgaccctg aggccttcac agagaatggg 660
gagtgggcca tccagcaccg accagccaag atgctcctgg acccagcggc gccagcccag 720
gaagcaggcc accagaaggt ggtgttctac ctgctcatcc agcgcaagcc cctcttctac 780
gtcatcaaca tcatcgcccc ctgtgtgctc atctcctctg tcgccatcct catccacttc 840
cttcctgcca aggctggggg ccagaagtgt accgtcgcca tcaacgtgct cctggcccag 900
actgtcttcc tcttccttgt ggccaagaag gtgcctgaaa cctcccaggc ggtgccactc 960
atcagcaagt acctgacctt cctcctggtg gtgaccatcc tcattgtcgt gaatgctgtg 1020
gttgtgctca atgtctcctt gcggtctcca cacacacact ccatggcccg aggggtgttc 1080
ctgaggctct tgccccagct gctgaggatg cacgttcgcc cgctggcccc ggcagctgtg 1140
caggacaccc agtcccggct acagaatggc tcctcgggat ggtcgatcac aactggggag 1200
gaggtggccc tctgcctgcc tcgcagtgaa ctcctcttcc agcagtggca gcggcaaggg 1260
ctggtggcgg cagcgctgga gaagctagag aaaggcccgg agttagggct gagccagttc 1320
tgtggcagcc tgaagcaggc tgccccagcc atccaggcct gtgtggaagc ctgcaacctc 1380
attgcctgtg cccggcacca gcagagtcac tttgacaatg ggaatgagga gtggttcctg 1440
gtgggccgag tgctggaccg cgtctgcttc ctggccatgc tctcgctctt catctgtggc 1500
acagctggca tcttcctcat ggcccactac aaccgggtgc cggccctgcc attccctgga 1560
gatccacgcc cctacctgcc ctcaccagac tga 1593

<210> 40
<211> 2121
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3116448CB1

<400> 40
gtacaaagga cctccagacc agagccagcc agcagcaaaa agagcatgga gctgaggagt 60
acagcagccc ccagagctga gggctacagc aacgtgggct tccagaatga agaaaacttt 120
cttgagaacg agaacacatc aggaaacaac tcaataagaa gcagagctgt gcaaagcagg 180
gagcacacaa acaccaaaca ggatgaagaa caggtcacag ttgagcagga ttctccaaga 240
aacagagaac acatggagga tgatgatgag gagatgcaac aaaaagggtg tttggaaagg 300
aggtatgaca cggtatgtgg tttctgtagg aaacacaaaa caactcttcg gcacatcatc 360
tggggcattt tattagcagg ttatctggtt atggtgattt cggcctgtgt gctgaacttt 420
cacagagccc ttcctctttt tgtgatcacc gtggctgcca tcttctttgt tgtctgggat 480
cacctgatgg ccaaatacga acatcgaatt gatgagatgc tgtctcctgg cagaaggctt 540
ctaaacagcc attggttctg gctgaagtgg gtgatctgga gctccctggt cctagcagtt 600
attttctggt tggcctttga cactgccaaa ttgggtcaac agcagctggt gtccttcggt 660
gggctcataa tgtacattgt cctgttattt ctattttcca agtacccaac cagagtttac 720
tggagacctg tcttatgggg aatcgggcta cagtttcttc ttgggctctt gattctaagg 780
actgaccctg gatttatagc ttttgattgg ttgggcagac aagttcagac ttttctggag 840
tacacagatg ctggtgcttc atttggcttt ggtgagaaat acaaagacca cttctttgga 900
tttaaggtcc tggcgatcgt ggtttcttc agcactgtga tgtccatgct gtactacctg 960
ggactgatgc agtggattat tagaaaggtt ggatggatca tgctagttac tacgggatca 1020
tctcctattg aatctgtagt tgcttctggc aatatatttg ttggacaaac ggagtctcca 1080
ctgctggtcc gaccatattt accttacatc accaagtctg aactccacgc catcatgacc 1140
gccgggttct ctaccattgc tggaagcgtg ctaggtgcat acatttcttt tggggttcca 1200
tcctcccact tgttaacagc gtcagttatg tcagcacctg cgtcattggc tgctgctaaa 1260
ctcttttggc ctgagacaga aaaacctaaa ataaccctca agaatgccat gaaaatggaa 1320
agtggtgatt cagggaatct tctagaagct gcaacacagg gagcatcctc ctccatctcc 1380
ctggtggcca acatcgctgt gaatctgatt gccttcctgg ccctgctgtc ttttatgaat 1440
tcagccctgt cctggtttgg aaacatgttt gactacccac agctgagttt tgagctaatc 1500
tgctcctaca tcttcatgcc cttttccttc atgatgggag tggaatggca ggacagcttt 1560

atggttgcca gactcatagg ttataagacc ttcttcaatg aatttgtggc ttatgagcac 1620
ctctcaaaat ggatccactt gaggaaagaa ggtggaccca aatttgtaaa cggtgtgcag 1680
caatatatat caattcgttc tgagataatc gccacttacg ctctctgtgg ttttgccaat 1740
atcgggtccc taggaatcgt gatcggcgga ctcacatcca tggctccttc cagaaagcgt 1800
gatatcgcct cgggggcagt gagagctctg attgcgggga ccgtggcctg cttcatgaca 1860
gcctgcatcg caggcatact ctccagcact cctgtggaca tcaactgcca tcacgtttta 1920
gagaatgcct tcaactccac tttccctgga aacacaacca aggtgatagc ttgttgccaa 1980
agtctgttga gcagcactgt tgccaagggt cctggtgaag tcatcccagg aggaaaccac 2040
agtctgtatt ctttgaaggg ctgctgcaca ttgttgaatc catcgacctt taactgcaat 2100
gggatctcta atacattttg a 2121

<210> 41
<211> 1225
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 622868CB1

<400> 41
aattcattgg catgaagtct cggacatggg cgtctgtcca tttgcattcc ttttttgcag 60
ttggaaccct gctggtggct ttgacaggat acttggtcag gacctggtgg ctttaccaga 120
tgatcctctc cacagtgact gtcccctta tcctgtgctg ttgggtgctc ccagagacac 180
ctttttggct tctctcagag ggacgatatg aagaagcaca aaaaatagtt gacatcatgg 240
ccaagtggaa cagggcaagc tcctgtaaac tgtcagaact tttatcactg gacctacaag 300
gtcctgttag taatagcccc actgaagttc agaagcacaa cctatcatat ctgttttata 360
actggagcat tacgaaaagg acacttaccg tttggctaat ctggttcact ggaagtttgg 420
gattctactc gttttccttg aattctgtta acttaggagg caatgaatac ttaaacctct 480
tcctcctggg tgtagtggaa attcccgcct acaccttcgt gtgcatcgcc acggacaagg 540
tcgggaggag aacagtcctg gcctactctc ttttctgcag tgcactggcc tgtggtgtcg 600
ttatggtgat cccccagaaa cattatattt tgggtgtggt gacagctatg gttggaaaat 660
ttgccatcgg ggcagcattt ggcctcattt atctttatac agctgagctg tatccaacca 720
ttgtaagatc gctggctgtg ggaagcggca gcatggtgtg tcgcctggcc agcatcctgg 780
cgccgttctc tgtggacctc agcagcattt ggatcttcat accacagttg tttgttggga 840
ctatggccct cctgagtgga gtgttaacac taaagcttcc agaaaccctt gggaaacggc 900
tagcaactac ttgggaggag gctgcaaaac tggagtcaga gaatgaaagc aagtcaagca 960
aattacttct cacaactaat aatagtgggc tggaaaaaac ggaagcgatt acccccaggg 1020
attctggtct tggtgaataa atgtgccatg cctgctgtct agcacctgaa atattattta 1080
ccctaatgcc tttgtattag aggaatctta ttctcatctc ccatatgttg tttgtatgtc 1140
tttttaataa attttgtaag aaaattttaa agcaaatatg ttataaaaga aataaaaact 1200
aagatgaaaa ttctcagttt taaaa 1225

<210> 42
<211> 2693
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476494CB1

<400> 42
tcctttgaga cagctgctct gagagaatgc aataagcagg gagcagccag caattcctcc 60
tagcagaggg cgactcgtgg gaggagttca gtttgccaag tattgtcatt tgttgagaga 120
aggtgtgtgc tcaaggagga gttttaacct ggaggatcat taactctttt agtcagctga 180

ggagctgcgg tggctcggcg agttggagtt catcctggaa gcgtctgcac gacaaggtca 240
gggatgaggt gtggaataac tttttcatgg gacactttga gaagggccag cacgctctgc 300
tcaatgaagg agaagagaat gagatggaga tatttggcta tcggactcaa ggctgccgga 360
aaagtctctg ccttgccgga tccatcttct catttggaat cctccccttg gtgttttact 420
ggagaccagc atggcacgta tgggcacatt gtgtcccatg ttccttgcaa gaagcagaca 480
ctgtgttgct gaggacaacg gtgagatgca tcaaagtgca gaaaataaga tatgtttgga 540
actacttaga aggacagttc cagaaaattg gttctttgga agactggctc agttctgcca 600
agatacatca aaaatttgga tcaggcttga caagagaaga acaggagatt aggaggttaa 660
tgtgtgggcc taatactatc gatgttgaag ttacaccaat ttggaaactg ctcatcaagg 720
aggttctaaa tccattttat atatttcaac tcttcagtgt ctgtttgtgg tttagtgaag 780
actataagga atatgctttt gccatcataa tcatgtccat aatttccata tctttgacag 840
tatatgatct cagagagcaa tctgtaaaac tccaccatct cgtcgagtca cataatagca 900
ttacggtctc tgtatgtggg agaaaagctg gagttcaaga gctggaatca cgcgtcctgg 960
tgcctggaga tttattaatt ttgacaggga acaaagtgct aatgccatgt gatgccgttc 1020
tgattgaagg cagctgtgtg gtggatgaag gcatgctgac aggagaaagt attccagtca 1080
ccaaaactcc gttacccaag atggatagct ctgtgccctg gaaaacacag agtgaagcgg 1140
attacaagcg gcatgtcctc ttctgtggaa cagaggttat ccaggccaag gcagcttgct 1200
ctgggaccgt gagagccgtg gtactgcaga ctggattcaa cactgcaaag ggagaccttg 1260
tgagatccat tctctaccct aagccagtga attttcagtt gtacagggat gccatcaggt 1320
tcctcctgtg ccttgtagga acagccacca ttgggatgat ctatactctg tgtgtctatg 1380
tgcttagtgg ggaacctcca gaggaggtgg tgaggaaagc ccttgacgtc atcacaattg 1440
cggttcctcc ggctctacct gctgctctga ccacaggcat tatctatgcc cagaggaggc 1500
tgaagaagag aggcatcttc tgcattagcc cccagaggat caacgtatgt ggacagttaa 1560
accttgtctg ctttgacaag acaggcacct taacaaggga cggcttggac ctctggggag 1620
tcgtgtcctg tgataggaat ggctttcagg aagttcacag ctttgcctca ggccaggctt 1680
tgccatgggg cccactgtgt gcagcgatgg ccagctgcca ctctctgatc cttcttgatg 1740
ggaccatcca gggagaccct ctggacctca aaatgtttga agccaccacc tgggaaatgg 1800
ctttttctgg ggacgatttc cacatcaagg gagtgccggc acatgccatg gtagttaagc 1860
cctgcagaac agccagccag gtcccagtgg aaggaattgc aatcctgcat cagttcccat 1920
tctcatcggc actgcaaaga atgacagtca ttgtccaaga gatgggaggt gaccgactgg 1980
cattcatgaa aggtgcacca gagagggtgg ccagcttttg ccaacctgag acagtaccca 2040
ctagttttgt tagcgaactt cagatttaca cgacacaggg cttccgagtc atagcactgg 2100
cctacaagaa gctggaaaat gaccatcacg ctactacctt gacgagggag acggtagaat 2160
cagacctgat atttctgggg ctgctgatct tggagaatcg attgaaggaa gagacaaaac 2220
ctgtcttgga agagctcatc tcagcccgga taaggactgt aatgatcaca ggtgacaatc 2280
ttcagactgc aataacagtg gccagaaaat ctggaatggt ttctgaaagc cagaaagtca 2340
ttctcattga ggcaaatgaa accaccgggt cctcatcagc atctatatct tggacgttag 2400
tagaagagaa gaaacacatt atgtatggga atcaggacaa ttacattaac atcagggatg 2460
aagtctctga taaaggcaga gaaggaagtt accatttgc cctaactgga aaatcctttc 2520
atgttataag tcaacatttc agcagcctac tgccaaagat attgatcaat gggaccatct 2580
ttgcaagaat gtctcctggg cagaagtcca gtctggtgga agaatttcag aaactggagt 2640
aggttctttg ccagtgcagg tggcatgaac tgcatggagg cataacagtc agg 2693

<210> 43
<211> 3569
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477260CB1

<400> 43
agaccagtgt tggaggatgg cttgcttggg gcccggtggg aagaagaacc ccctcgggtg 60
gtaagctgaa gttgggtcag agtgcttctc acttctctct tcagttctgg ttcttgctac 120
tgccctggct tcgaccattc ccccatcatt cttcactgcc agctgcaaga ccctggatct 180

gaatgcagac taaatctttt catctctttt atcttaaaga gtatctcgcc cacctctgat 240
tcatggttac aggaggccag catcacccag gagctggact tagttttaca gaattagaaa 300
atacttttcc cttgtgcttg cctcctactc catttctgtt ggccttgtgg tcctcctgcc 360
ttccatggga cactcagcag acctgctgcc cctcttttgc agggtcccca gctgctgagc 420
agctccagga catcctgggg gaggaagatg aggctcccaa ccccaccctc tttacagaga 480
tggatactct gcagcatgac ggagaccaga tggagtggaa ggagtcagcc aggtggataa 540
agtttgaaga aaaggtagag gaaggcggcg aacgctggag caagccccac gtgtccacac 600
tatccctgca cagcctcttc gagctccgta cctgcctgca gacggggacg gtgctgctgg 660
atttggacag tggctcctta ccacagatca tagatgatgt cattgagaag cagattgagg 720
atggtctcct gcggccagag ctccgggaga gggtcagtta cgtcctcctg aggaggcacc 780
gccaccaaac caagaagccc atccaccgct ccttagctga cattgggaag tcagtctcca 840
ccacaaatcg cagtcctgcc cggagccctg gtgctggccc gagtctacac cactccacgg 900
aagacctgcg gatgcggcag agtgcaaatt acggacgtct gtgtcatgcc cagagcagaa 960
gcatgaatga catttctctc accccaaaca cagaccagcg gaaaaacaaa ttcatgaaga 1020
agatccccaa ggactcagaa gcgtccaacg tgctcgtggg cgaggtggac ttcctagacc 1080
agccattcat cgcgttcgtg cgcctcatcc agtcggccat gctgggagga gtgaccgagg 1140
tgcctgtccc caccagattt ctgtttatac tactgggacc ttctgggaga gcaaaatcct 1200
acaatgaaat tggccgtgcc attgcaaccc tcatggtaga tgatctcttc agtgacgtgg 1260
cctacaaagc ccgcaatcgg gaagatctga tcgcaggaat tgatgaattt ctggatgagg 1320
tcatcgtcct tcctcctgga gaatgggacc caaatatccg gattgagccc cccaagaagg 1380
tgccctctgc tgacaagagg aaatctctgt tctccctagc agagctgggc cagatgaatg 1440
gctctgtggg aggaggcggc ggagctcctg gaggaggcaa tggaggtggt ggtggtggtg 1500
gcagtggcgg cggggctggc agtggcgggg ccggcggaac aagcagcggg gatgatggag 1560
agatgccagc catgcatgaa atcggggagg aacttatctg gacaggaagg ttcttcggtg 1620
gcctgtgtct ggatatcaag aggaagttgc cctggttccc aagtgacttc tatgatggct 1680
tccacattca gtccatctct gccatcctat tcatctacct cggctgtatc accaacgcga 1740
tcacctttgg tgggcttctg gggggatgcca ccgacaatta tcagggagtg atggagagct 1800
tcctgggcac tgccatggct ggctccttgt tctgcctctt ctcgggacag cctctcatca 1860
ttctcagcag cacgggggccc atcctcatct ttgagaagct cctcttcgac ttcagcaaag 1920
gcaatggcct ggactacatg gagttccgcc tctggattgg cctacactca gctgtccagt 1980
gccttatcct agtggccaca gatgccagct ttatcatcaa atatatcacc cgcttcaccg 2040
aggagggctt ctccaccctt atcagcttca tcttcatcta cgatgccatc aagaagatga 2100
tcggtgcctt caagtactac cctatcaata tggacttcaa gccaaacttc atcactacct 2160
acaagtgcga gtgtgtcgcc cctgacacag gtgacctgaa tacaaccgtg ttcaatgctt 2220
cagccccatt ggcaccagac accaacgctt ctctgtacaa cctccttaac ctcacagcgt 2280
tggactggtc cctgctgagc aagaaggagt gtctgagcta cggtgggcgc ctgcttggga 2340
attcctgcaa gtttatccca gacctggcgc tcatgtcctt catccttttc tttgggacat 2400
actccatgac cctgaccctg aagaagttca aattcagccg ctattttcct accaaggtcc 2460
gggccctggt ggctgacttt tccattgttt tctccatcct gatgttctgt ggaatcgatg 2520
cctgttttgg cctagaaact cccaagctgc atgtgcccag tgtcatcaag ccaacgcggc 2580
ctgaccgagg ctggttcgtg gcccccttttg ggaagaaccc gtggtgggta tacccagcaa 2640
gcatcctgcc cgccctgctg gtgaccatcc tgatcttcat ggaccagcag atcactgccg 2700
tcattgtcaa ccggaaggag aacaaactga agaaggctgc cggctaccat ctggacctgt 2760
tctgggtggg catcctcatg gctttgtgct cctttatggg gctcccctgg tacgtggctg 2820
ccacggtcat ctccatcgcc cacatcgaca gcctcaagat ggagacagag accagtgccc 2880
ctggggagca gccccagttt ctgggagtca gggaacagag agtaaccggc atcatcgtct 2940
tcatcctgac gggaatctct gtcttcctgg ctcccatcct aaagtgtatc cccctgccgg 3000
tgctgtacgg agtcttcctc tacatgggcg tggcctccct gaatggcatc cagttctggg 3060
aacgctgcaa gctcttcctg atgccagcca agcaccagcc ggaccatgcc ttcctgcggc 3120
acgtgccgct gcgccggatc cacctcttca ccctggtgca gatcctctgc ctggcggtgc 3180
tctggatcct caaatccacg gtggctgcca tcatcttccc ggtcatgatc ctgggcctca 3240
tcatcgttcg aaggcttctg gatttcatct tttcccagca cgacctggcc tggattgaca 3300
acatcctccc agagaaggaa aaaaaggaga cagacaagaa gaggaagaga aaaaaagggg 3360
cccacgagga ctgtgatgag gaggaaaaag atcttccagt tggagttact cactctgatt 3420
cttccttcag tgacacagaa cttgaccgaa gctactcacg gaacccagtg ttcatggtgc 3480
cacaggtgaa gatagagatg gagtcagact atgacttcac agacatggat aaataccgaa 3540

gagaaactga cagtgagacc accctctag 3569

<210> 44
<211> 3920
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1963058CB1

<400> 44
cggacgcggc ggacgtgggt gagggcgcgg ccgtaagaga gcgggacgcg gggtgcccgg 60
cgcgtggtgg gggtccccgg cgcctgcccc cacggcaccc aagaaggcct ggccagggta 120
ccctccgcgg agcccggggg tggggggcgc ggggccggcg ccgcgatggg cccgggaccc 180
ccagcggccg gagcggcgcc gtccccgcgg ccgctgtccc tggtggcgcg gctgagctac 240
gccgtgggcc acttcctcaa cgacctgtgc gcgtccatgt ggttcaccta cctgctgctc 300
tacctgcact cggtgcgcgc ctacagctcc cgcggcgcgg ggctgctgct gctgctgggc 360
caggtggccg acggctgtg cacaccgctc gtgggctacg aggccgaccg cgccgccagc 420
tgctgcgccc gctacggccc gcgcaaggcc tggcacctgg tcggcaccgt ctgcgtcctg 480
ctgtccttcc ccttcatctt cagcccctgc ctgggctgtg gggcggccac gcccgagtgg 540
gctgccctcc tctactacgg cccgttcatc gtgatcttcc agtttggctg ggcctccaca 600
cagatctccc acctcagcct catcccggag ctcgtcacca acgaccatga gaaggtggag 660
ctcacggcac tcaggtatgc gttcaccgtg gtggccaaca tcaccgtcta cggcgccgcc 720
tggctcctgc tgcacctgca gggctcgtcg cgggtggagc ccacccaaga catcagcatc 780
agcgaccagc tgggggggcca ggacgtgccc gtgttccgga acctgtccct gctggtggtg 840
ggtgtcggcg ccgtgttctc actgctattc cacctgggca cccgggagag gcgccggccg 900
catgcggagg agccaggcga gcacaccccc ctgttggccc ctgccacggc ccagcccctg 960
ctgctctgga agcactggct ccgggagccg gctttctacc aggtgggcat actgtacatg 1020
accaccaggc tcatcgtgaa cctgtcccag acctacatgg ccatgtacct cacctactcg 1080
ctccacctgc ccaagaagtt catcgcgacc attcccctgg tgatgtacct cagcggcttc 1140
ttgtcctcct tcctcatgaa gcccatcaac aagtgcattg ggaggaacat gacctacttc 1200
tcaggcctcc tggtgatcct ggcctttgcc gcctgggtgg cgctggcgga gggactgggt 1260
gtggccgtgt acgcagcggc tgtgctgctg ggtgctggct gtgccaccat cctcgtcacc 1320
tcgctggcca tgacggccga cctcatcggt ccccacacga acagcggagc gttcgtgtac 1380
ggctccatga gcttcttgga taaggtggcc aatgggctgg cagtcatggc catccagagc 1440
ctgcaccctt gcccctcaga gctctgctgc agggcctgcg tgagctttta ccactgggcg 1500
atggtggctg tgacgggcgg cgtgggcgtg gccgctgccc tgtgtctctg tagcctcctg 1560
ctgtggccga cccgcctgcg acgctgggac cgtgatgccc ggccctgact cctgacagcc 1620
tcctgcacct gtgcaaggga actgtgggga cgcacgagga tgccccccag ggccttgggg 1680
aaaagccccc actgcccctc actcttctct ggacccccac cctccatcct cacccagctc 1740
ccgggggtgg ggtcgggtga gggcagcagg gatgcccgcc agggacttgc aaggaccccc 1800
tgggttttga gggtgtccca ttctcaactc taatccatcc cagccctctg gaggatttgg 1860
ggtgcccctc tcggcaggga acaggaagta ggaatcccag aagggtctgg gggaacccta 1920
accctgagct cagtccagtt cacccctcac ctccagcctg gggtctcca gacactgcca 1980
gggcccccctc aggacggctg gagcctggag gagacagcca cggggtggtg ggctgggcct 2040
ggaccccacc gtggtgggca gcagggctgc ccggcaggct tggtggactc tgctggcagc 2100
aaataaagag atgacggcag cctggctcct gtctgcctgc ggggggggctc tgggcagggg 2160
tagcctgggc atctcagccc tgccctggtt gtgggcggcc agcgagccca gtgtctgcct 2220
ctgtcccgag cctctggtcc cctgggacta ggttagtgcc ccctcatctg ggtgcagaga 2280
cagtgggtgc atcctggtag catgccttta tcggggagtg ggtgtgaggg aaggcgggga 2340
ccgctggcag gtggaggggc agtatggttc caggacccac tcccggtagt tctgggtggt 2400
gccgggcggg cgctggggtg ccgacaggga gggcacgtag tctgatgccc tccacagtgg 2460
ctccaccccg taccggttcc tgttgagcac tgtaggtggg actcgggtca ccatgtgccc 2520
ccacctcctc gccgttggcc agcaaggggc tcctggatcg ccccgggcag tttcaccctg 2580
gcctaggtgg ccttgtcccc ctggcctccc aaggacccac cctgcaccta gcctcaccgt 2640

attccttgcc ccggattggc ctgtctttcc acagcgcgct cccccaccgg gtgctggggg 2700
cctggtactg ggcagggacg atggggtcat gccaggcggt ctcccgcagg tgctgggtgt 2760
aggctgcggt ggggcggggg ctggcggtca ttcctgtccc cctctggcag gcccgctgcc 2820
cagggcgggg gggggggcac tcacccgatg gcatgctgca ctcacggtgt tggtagcagc 2880
tgtgccagcg gttgtaggcc tcgcggaagg ggctggccgg ggcccgtggc aggttgtacc 2940
aggcttccca ggcgtccgag ttggtcaggc ctgtgtacca cagctggccg gctgcgtccc 3000
gtcccatggg cgtgtacttc cagcgtgtgg cctgcctgat ggctggtggc cagcgggaac 3060
cctccaggga caggtagtca tcgctgaggg tgggcagggg gcagagactg agcccatgtc 3120
tacagcgagt gctttgaccc ctttgcgatg tctgccaggg tggatgatgt agaggcctgg 3180
cccacggcgt ggggtctccc tccctcgcca cttggagtct gtccttcagc cctgtacccc 3240
tcaccccaga gtgggtgctt gaggagagag gctgactccc cctctcccca cacatcgcac 3300
cccaagcacc caagtcagca ctaaaccttt ctgttctcag cttttcttgc ctggagaaga 3360
gggaggggag aggacaaggg ccctggctac tcctggattc ctacagtcct tgtccagcct 3420
ccaagaccca caagtccctt cctctgggaa gcccccctgg cctggaggtg caccaggaag 3480
aagtggtctg gggctggcac taagccatgg cccagggaag actgggggac ccactaggcc 3540
aggtgtgtgg ctcacgcttg taaacccagc actttgggag gctgaggcag gtggatcact 3600
tgaggtcagg agttcgagac cagcctggcc agcatggtaa aaccccatct ctactaaaaa 3660
tacgaaaatt aagccaggca tggtgtgggg gcgggggggca cctgtaatcc cagctactca 3720
ggaggctgag gcaggagaat cgcttgaacc caggaagtgg agtttgcagt aagctgagat 3780
cgtgcccttt gactccagcc ctgggaaaag agtgagactc cgtcctcaaa aaccaaaggg 3840
ccaggagact caaagaatgt cttatgcttt gaaccttgct ccttggaata atgtcccagg 3900
gaagtcatcc cagaaaacaa 3920

<210> 45
<211> 1361
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2395967CB1

<400> 45
ctggaagcat gtcggagttt tggttaattt ctgcccctgg cgataaggaa aatttgcaag 60
ctctggagag gatgaatact gtaacctcca agtccaacct gtcttataat accaaattcg 120
ctattcctga cttcaaggtg gggaccttgg attccctggt tggcctctct gatgagttgg 180
ggaaactcga cacctttgct gaaagcctca taaggagaat ggctcagagc gtggtggaag 240
tcatggagga ctcaaagggg aaggtccagg agcacctcct ggcaaacgga gttgacttaa 300
catcctttgt gacccacttt gaatgggaca tggccaaata tcctgtcaag cagccgctcg 360
tgagtgtggt ggacacaata gccaagcaac tggcgcagat cgagatggac ctgaagtccc 420
gaacggccgc ctacaacact ctgaagacaa acctggagaa cctggaaaag aaatccatgg 480
ggaacctctt cacccggaca ctgagtgata ttgtgagcaa agaggacttc gtgctggatt 540
ctgaatatct cgtcacactt ctggtcatcg tccccaaacc aaactactca caatggcaaa 600
aaacctacga atctctctca gacatggtag tccctcgatc aaccaaactc attactgagg 660
acaaggaagg gggccttttc actgtgactc tgtttcgaaa agtgattgaa gatttcaaaa 720
ccaaggccaa agaaaacaag ttcactgttc gtgaatttta ctatgatgag aaggaaattg 780
aaagggaaag ggaggagatg gccagattgc tgtctgataa gaagcaacag tatggccccc 840
tgctgcgctg gctcaaggtg aacttcagtg aagccttcat tgcctggatc cacatcaagg 900
ccctgagagt gtttgtggag tccgtgctca ggtatggact accagtgaac ttccaggcag 960
tgctcctgca gccgcataag aagtcatcca ccaagcgttt aagagaggtt ctaaactctg 1020
tcttccgaca tctggatgaa gtagccgcta caagtatact ggatgcatct gtggagatcc 1080
cgggactgca actcaataac caagactatt ttccttatgt ctacttccat attgacctta 1140
gtcttcttga ctagaaaggc cagctggcac ctctgtctca tgttcgtgca gattattaca 1200
gacacctctt tcctttagcc agagaatggt tcaaatgtct tacagaacta agatcttttt 1260
cagagaaatt gctcacaaaa gttagtgaca gttgtattta ttttttttaag ttacaataaa 1320
atgctctcaa gtcctttgaa tgttccaaca aattcaaaaa a 1361

<210> 46
<211> 1867
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3586648CB1

<400> 46
cagaattagc cggtatagga atgaacgagc atgaagattt gaaattgctc cgattggaag 60
gaagcccagg ttaggtttgg gcacctccaa acgcacccgt tttaaagcca cctggactga 120
ggcgtcgagc tttcagctcc accaaacgct cacctggcct ggcagcgagc ggcggaagag 180
cccgggagcc cctcacagag cgcaccgagc cgggcggaga gctgagccgc aggcacccgc 240
gtctccagga tgataggcga cattgcaaca aatctctaca cccagcagct caggggggctc 300
caagcagagc agcaagttcg aggatccggg cgtggagccg agtgaggccg cagcccagcg 360
ggcctcgggc gaaaaatctt ggaaaatgta taccagtcat gaagatattg ggtatgattt 420
tgaagatggc cccaaagaca aaaagacact gaagccccac ccaaacattg atggcggatg 480
ggcttggatg atggtgctct cctctttctt tgtgcacatc ctcatcatgg gctcccagat 540
ggccctgggt gtcctcaacg tggaatggct ggaagaattc caccagagcc gcggcctgac 600
cgcctgggtc agctccctca gcatgggcat caccttgata gtgggccctt tcatcggctt 660
gttcattaac acctgtgggt gccgccagac tgcgatcatt ggagggctcg tcaactccct 720
gggctgggtg ttgagtgcct atgctgcaaa cgtgcattat ctcttcatta cttttggagt 780
cgcagctggc ctgggcagcg ggatggccta cctgccagcg gtggtcatgg tgggcaggta 840
tttccagaag agacgcgccc tcgcccaggg cctcagcacc acggggaccg gattcggtac 900
gttcctaatg actgtgctgc tgaagtacct gtgcgcagag tacggctgga ggaatgccat 960
gttgatccaa ggtgccgttt ccctaaacct gtgtgtttgt ggggcgctca tgaggcccct 1020
ctctcctggt aaaaacccaa acgacccagg agagaaagat gtgcgtggcc tgccagcgca 1080
ctccacagaa tctgtgaagt caactggaca gcagggaaga acagaagaga aggatggtgg 1140
gctcgggaac gaggagaccc tctgcgacct gcaagcccag gagtgccccg atcaggccgg 1200
gcacaggaag aacatgtgtg ccctccggat tctgaagact gtcagctggc tcaccatgag 1260
agtcaggaag ggcttcgagg actggtattc gggctacttt gggacagcct ctctatttac 1320
aaatcgaatg tttgtagcct ttattttctg ggctttgttt gcatacagca gctttgtcat 1380
ccccttcatt cacctcccag aaatcgtcaa tttgtataac ttatcggagc aaaacgacgt 1440
tttccctctg acgtcaatta tagcaatagt tcacatcttt ggaaaagtga tcctgggcgt 1500
catagccgac ttgccttgca ttagtgtttg gaatgtcttc ctgttggcca acttcacccct 1560
tgtcctcagt atttttattc tgccgttgat gcacacgtac gctggcctgg cggtcatctg 1620
tgcgctgata gggttttcca gtggttattt ctccctaatg cccgtagtga ctgaagactt 1680
ggttggcatt gaacacctgg ccaatgccta cggcatcatc atctgtgcta atggcatctc 1740
tgcattgctg ggaccacctt ttgcaggtaa actctctgag gttttaagag ctcagagtgc 1800
atgtacatat ggtgcgttat gttataaagt cccagattaa gaaacaaaaa aaaaaaaaaa 1860
agatcgg 1867

<210> 47
<211> 2211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473396CB1

<400> 47
atgcagaata ttaccaaaga atttggaaca ttcaaggcaa atgacaacat caatttacaa 60
gtaaaggcag gagagattca tgcgttgctt ggagaaaacg gtgctggcaa atctacattg 120
atgaacgtgc tttccggatt attagagccg acatcaggga aaattttgat gcgtgggaaa 180

gaagtacaga tcacaagccc gacaaaagcc aatcaattag ggattgggat ggtccatcag 240
cactttatgc ttgttgatgc ctttactgta acagaaaaca tcgtgttggg aagcgaacct 300
agtcgtgcag ggatgcttga ccataaaaaa gcgcgaaaag agatccaaaa agtttctgaa 360
caatatggat tatcagtcaa cccggatgct tatgttcgtg atatttcagt tgggatggaa 420
caacgggtag aaattttaaa aacactttac cgaggagcag atgtactgat tttgatgag 480
ccgacagctg tattgacccc tcaggaaatt gatgaattaa tcgtgatcat gaaggaatta 540
gtcaaagaag gcaagtcaat cattttgatt acgcataagt tagatgaaat caaagcagta 600
gctgaccgtt gtacagttat ccgccgtgga aaaggaatcg gtacagtcaa cgttaaagac 660
gttacctcac agcaattagc tgatatgatg gtcggaagag cggtttcatt caaaacgatg 720
aaaaaagaag cgaagcctca agaagtcgtt ttgtctattg aaaatctagt ggtaaaagaa 780
aatcgtggat tagaagccgt gaaaaacctg aacttagagg ttcgtgctgg cgaagtactt 840
ggtatcgctg gaatcgatgg aaacgggcag tcggagttga tccaagcttt gactggtttg 900
cgaaaggcag aaagcggaca tatcaagcta aaaggggaag acatcaccaa taaaaaacct 960
cgaaagatca ctgaacatgg tgtaggacat gtgccagaag accgtcataa atacgggttg 1020
gtcctagata tgacattgtc tgaaaacatt gccctgcaaa cgtatcatca aaaaccttac 1080
agtaaaaacg gtatgctgaa ttattcagtg ataaatgaac atgccagaga attgatcgaa 1140
gaatatgatg ttcgaacaac gaatgaactt gttcctgcaa aagctttatc aggcggaaat 1200
cagcaaaaag caatcatcgc tcggatagtc gaccgagatc ctgatctgtt gatcgttgca 1260
aatccaactc gtgggctgga tgtaggagaa tttgtagcag tcacaggtgt gtctggttct 1320
ggaaagagta cattggtcaa tagtatctta aagaaatcgt tagcgcaaaa attaaataag 1380
aattctgcta agccaggtaa attcaagaca atttccggct acgaaagtat cgaaaagatc 1440
atcgatatcg atcaaagccc aatcggccgg acgccgagaa gtaatccagc gacttataca 1500
agtgtatttg atgatatccg tgggttattt gctcaaacga acgaggcaaa aatgcggggt 1560
tataagaaag ggcgttttag tttcaacgta aaaggcggtc gttgtgaagc ttgtcgcggg 1620
gatggaatta ttaagatcga aatgcacttt ttgcctgatg tctatgttcc ttgtgaagta 1680
tgtcatggca aacgatataa ctctgaaaca ttagaagtgc attacaaagg aaaaagcatt 1740
gctgatattt tggaaatgac agtagaagat gctgtagaat tcttcaagca cattccaaag 1800
attcatcgca aactgcaaac gattgttgat gttggcttag gttatgtgac tatggggcaa 1860
ccagcaacga cattgtccgg tggtgaggca caacggatga aacttgccag tgaattgcac 1920
aaaatctcta atggaaagaa tttctatata ctagatgaac caacgacagg acttcatagc 1980
gatgacatcg cccgcttgtt gcatgtatta caaagattag tagatgctgg taacacagtt 2040
ttagtgattg aacacaatct agatgtaatc aaaacagcag attatatcat tgatttagga 2100
ccagaaggtg gagaaggtgg aggaacgatc cttacgactg gaacaccaga agaaatcatt 2160
aacgtaaaag aaagttatac aggtcactat ttgaaaaaaa taatggtata a 2211

<210> 48
<211> 1446
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476283CB1

<400> 48
tggctgggag aattgagcta gtgcagcaca cgtaaaaaag cgattccgat gggtcctttg 60
aaagcttttc tcttctcccc ttttcttctg cggagtcaaa gtagaggggt gaggttggtc 120
ttcttgttac tgaccctgca tttgggaaac tgtgttgata aggcagatga tgaagatgat 180
gaggatttaa aggtgaacaa aacctgggtc ttggccccaa aaattcatga aggagatatc 240
acacaaattc tgaattcatt gcttcaaggc tatgacaata aacttcgtcc agatatagga 300
gtgaggccca cagtaattga aactgatgtt tatgtaaaca gcattggacc agttgatcca 360
attaatatgg aatatacaat agatataatt tttgcccaaa cctggtttga cagtcgttta 420
aaattcaata gtaccatgaa agtgcttatg cttaacagta atatggttgg aaaaatttgg 480
attcctgaca ctttcttcag aaactcaaga aaatctgatg ctcactggat aacaactcct 540
aatcgtctgc ttcgaatttg gaatgatgga cgagttctgt atactctaag attgacaatt 600
aatgcagaat gttatcttca gcttcataac tttcccatgg atgaacattc ctgtccactg 660

gaattttcaa gcgatggata ccctaaaaat gaaattgagt ataagtggaa aaagccctcc 720
gtagaagtgg ctgatcctaa atactggaga ttatatcagt ttgcatttgt agggttacgg 780
aactcaactg aaatcactca cacgatctct ggggattatg ttatcatgac aatttttttt 840
gacctgagca gaagaatggg atatttcact attcagacct acattccatg cattctgaca 900
gttgttcttt cttgggtgtc tttttggatc aataaagatg cagtgcctgc aagaacatcg 960
ttgggtatca ctacagttct gactatgaca accctgagta caattgccag gaagtcttta 1020
cctaaggttt cttatgtgac tgcgatggat ctctttgttt ctgtttgttt catttttgtt 1080
tttgcagcct tgatggaata tggaaccttg cattatttta ccagcaacca aaaaggaaag 1140
actgctacta aagacagaaa gctaaaaaat aaagcctcga tgactcctgg tctccatcct 1200
ggatccactc tgattccaat gaataatatt tctgtgccgc aagaagatga ttatgggtat 1260
cagtgtttgg agggcaaaga ttgtgccagc ttcttctgtt gctttgaaga ctgcagaaca 1320
ggatcttgga gggaaggaag gatacacata cgcattgcca aaattgactc ttattctaga 1380
atatttttcc caaccgcttt tgccctgttc aacttggttt attgggttgg ctatctttac 1440
ttataa 1446

<210> 49
<211> 1332
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477105CB1

<400> 49
ttcggctcga gggaccccag gccgggccgg gccgagaggc tgccatgggc tccgtgggga 60
gccagcgcct tgaggagccc agcgtggcag gcacaccaga cccgggcgta gtgatgagct 120
tcaccttcga cagtcaccag ctggaggagg cggcggaggc ggctcagggc cagggcctta 180
gggccagggg cgtcccagct ttcacggata ctacattgga cgagccagtg cccgatgacc 240
gttatcacgc catctacttt gcgatgctgc tggctggcgt gggcttcctg ctgccataca 300
acagcttcat cacggacgtg gactacctgc atcacaagta cccagggacc tccatcgtgt 360
ttgacatgag cctcacctac atcttggtgg cactggcagc tgtcctcctg aacaacgtcc 420
tggtggagag actgaccctg cacaccagga tcaccgcagg ctacctctta gccttgggcc 480
ctctcctttt tatcagcatc tgcgacgtgt ggctgcagct cttctctcgg gaccaggcct 540
acgccatcaa cctggccgct gtgggcaccg tggccttcgg ctgcacagtg cagcaatcca 600
gcttctacgg gcaccgcctg gcccagcctc caccagggac ccctcctcat gaactctgga 660
gccctgagag gagaggggca gccccccacc ttgtcaccct cagggcttcc ccttctgtcc 720
tcattcttag agactgcttc tcccaaacat aacgcgttag ccatgaagga gtcggagccc 780
tgggtccgaa tggacccgcc tgcggtctgc atcagcctct gggaaaccac agcagtgatg 840
ccagctgggc acgtcaggac ctccccacac acccacacga tgccacaggt caggggggctg 900
tgcctgacta gggagccctc ccattgcctt cctggcccgg gatagaagag gggaggtaag 960
tctgggggct acgaagccgg gcccccacac cctggctgaa gtcagcttga cctaggtctt 1020
gaccctcatc cagcaaggga ctcgacagac ccaagggtcc ctggaacgta gggaggggct 1080
gggggtcact ccagcccggg cctcccagaa caccaggccc gtgtgggtgg caccctgagg 1140
tcaggggatc ctaagggtgt ccttccagag acggtgtttc caggggggagg accgccccccg 1200
cttccagatc cccggccccg gctgtgactg ccctgtttca cccctgctgt gtcccatccc 1260
ccgtctgtcc actaactgta ccgcaccggc cattaaaaga tgaaggcaga ccgctggaaa 1320
aaaaaaaaaa aa 1332

<210> 50
<211> 2298
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7482079CB1

<400> 50
atgctcaaac agagtgagag gagacggtcc tggagctaca ggccctggaa cacgacggag 60
aatgagggca gccaacaccg caggagcatt tgctccctgg gtgcccgttc cggctcccag 120
gccagcatcc acggctggac agagggcaac tataactact acatcgagga agacgaagac 180
ggcgaggagg aggaccagtg gaaggacgac ctggcagaag aggaccagca ggcaggggag 240
gtcaccaccg ccaagcccga gggccccagc gaccctccgg ccctgctgtc cacgctgaat 300
gtgaacgtgg gtggccacag ctaccagctg gactactgcg agctggccgg cttccccaag 360
acgcgcctag gtcgcctggc cacctccacc agccgcagcc gccagctaag cctgtgcgac 420
gactacgagg agcagacaga cgaatacttc ttcgaccgcg acccggccgt cttccagctg 480
gtctacaatt tctacctgtc cggggtgctg ctggtgctcg acgggctgtg tccgcgccgc 540
ttcctggagg agctgggcta ctggggcgtg cggctcaagt acacgccacg ctgctgccgc 600
atctgcttcg aggagcggcg cgacgagctg agcgaacggc tcaagatcca gcacgagctg 660
cgcgcgcagg cgcaggtcga ggaggcggag gaactcttcc gcgacatgcg cttctacggc 720
ccgcagcggc gccgcctctg gaacctcatg gagaagccat tctcctcggt ggccgccaag 780
gccatcgggg tggcctccag caccttcgtg ctcgtctccg tggtggcgct ggcgctcaac 840
accgtggagg agatgcagca gcactcgggg cagggcgagg gcggcccaga cctgcggccc 900
atcctggagc acgtggagat gctgtgcatg ggcttcttca cgctcgagta cctgctgcgc 960
ctagcctcca cgcccgacct gaggcgcttc gcgcgcagcg ccctcaacct ggtggacctg 1020
gtggccatcc tgccgctcta ccttcagctg ctgctcgagt gcttcacggg cgagggccac 1080
caacgcggcc agacggtggg cagcgtgggt aaggtgggtc aggtgttgcg cgtcatgcgc 1140
ctcatgcgca tcttccgcat cctcaagctg gcgcgccact ccaccggact gcgtgccttc 1200
ggcttcacgc tgcgccagtg ctaccagcag gtgggctgcc tgctgctctt catcgccatg 1260
ggcatcttca ctttctctgc ggctgtctac tctgtggagc acgatgtgcc cagcaccaac 1320
ttcactacca tcccccactc ctggtggtgg gccgcggtga gtacctttgc cctgggcttt 1380
cccatcctct tccccagccc agtgagctgc tcctccctcc cctggttatc agccaccagg 1440
ctttggcttc tgatcctcgt cttccccccc acccccaatc gccgcataca gctaacaaaa 1500
cggcgatgga tgtcaaaagt ggtggaaaga gaactcagca gatcagtaaa ctccagcagc 1560
cacatgtcga tggctgtggc aaagaacaag agagagaatg caagccccat catgcaaaca 1620
cttcataagt ttcttttcat ggcatttgct cagcccattg gccagagtaa gtcacatggc 1680
caagctgcaa gtcaaagggc agggcaggtg agcatctcca ccgtgggcta cggagacatg 1740
tacccagaga cccacctggg caggtttttt gccttcctct gcattgcttt tgggatcatt 1800
ctcaacggga tgcccatttc catcctctac aacaagtttt ctgattacta cagcaagctg 1860
aaggcttatg agtataccac catacgcagg gagaggggag aggtgaactt catgcagaga 1920
gccagaaaga agatagctga gtgtttgctt ggaagcaacc cacagctcac cccaagacaa 1980
gagaattagt attttatagg acatgtggct ggtagattcc atgaacttca aggcttcatt 2040
gctctttttt taatcattat gattggcagc aaaaggaaat gtgaagcaga catacacaaa 2100
ggccatttcg ttcacaaagt actgcctcta gaaatactca ttttggccca aactcagaat 2160
gtctcatagt tgctctgtgt tgtgtgaaac atctgacctt ctcaatgacg ttgatattga 2220
aaacctgagg ggagcaacag cttagatttt tcttgtagct tctcgtggca tctagctcaa 2280
taaatatttt tggacttg 2298

<210> 51
<211> 2250
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55145506CB1

<400> 51
agaaacagat ctctcggatc aataagcatg aatgacgaag actacagcac catctatgac 60
acaatccaaa atgagaggac gtatgaggtt ccagaccagc cagaagaaaa tgaaagtccc 120
cattatgatg atgtccatga gtacttaagg ccagaaaatg atttatatgc cactcagctg 180

```
aatacccatg agtatgattt tgtgtcagtc tataccatta agggtgaaga gaccagcttg 240
gcctctgtcc agtcagaaga cagaggctac ctcctgcctg atgagatata ctctgaactc 300
caggaggctc atccaggtga gccccaggag gacaggggca tctcaatgga agggttatat 360
tcatcagccc aggaccagca actctgcgca gcagaactcc aggagaatgg gagtgtgatg 420
aaggaagatc tgccttctcc ttcaagcttc accattcagc acagtaaggc cttctctacc 480
accaagtatt cctgctattc tgatgctgaa ggtttggaag aaaaggaggg agctcacatg 540
aaccctgaga tttacctctt tgtgaaggct ggaatcgatg gagaaagcat cggcaactgt 600
cctttctctc agcgcctctt catgatcctc tggctgaaag gagtcgtgtt caatgtcacc 660
actgtggatc tgaaaagaaa gccagctgac ctgcacaacc tagcccccgg cacgcacccg 720
cccttcctga ccttcaacgg ggacgtgaag acagacgtca ataagatcga ggagttcctg 780
gaggagacct tgacccctga aaagtacccc aaactggctg caaaacaccg ggaatccaac 840
acagcgggca tcgacatctt ttccaagttt tctgcctaca tcaaaaatac caagcagcag 900
aacaatgctg ctcttgaaag aggcctaacc aaggctctaa agaaattgga tgactacctg 960
aacacccctc taccagagga gattgacgcc aacacttgtg gggaagacaa ggggtcccgg 1020
cgcaagttcc tggatgggga tgagctgacc ctggctgact gcaatctgtt gcccaagctc 1080
catgtggtca agacccacct tctcacttcc tccagcaact tcctaaggaa caagtaccac 1140
tgaaagggat gatataattc cagctcagtc acactgtgtc agagtgatac aatgcaaaga 1200
tcaggagacc cgagttccgg tcctgtattt gctgccaact agcagcatga gctgaggcac 1260
atcatttaat ctttttggaa ttcatttttc tcatgcctag aagaacagaa gtggattgta 1320
ttccttcttg ccttcttttc ctttcttctt tccctccttc tttccttttc tcttgctcaa 1380
acatgtattc actaccactc aaaaaccatt tgttgaacaa agcaaacaaa tgaatctccc 1440
aagccttggg cttcatcctg tgatttcctc aattcccacc tgccttaaat tactcagtga 1500
agccctgtcc ttggagaaaa ttcagtgggt ggttaaccca gagaagctgg agatcaaaaa 1560
gaagatggcc aatgaaagaa caaaggccag cccttggccc ctatctcttt ggatttctgc 1620
tgatccagct tatcagatcc cagaaacctg gcaaacctct aaagttcaca aagagcgaag 1680
gggaagccaa gtcaggcctc cagtttggct tcggatgcca aaacttaatc tgggctgtgg 1740
gagctaactg ttttcatatg aaagagcaaa ttcagaacat gagcatggaa gtccctgcga 1800
acgtcagatc tccgtgtgca tccttacccc cttgctgctt tcatgctcac tctcctcttg 1860
cgtggctcgc tttcaggttt atctccatcc ctggaagcag agttgctctg gcccaggctc 1920
tccatgagag tttggcttga acattcattg tctggccccc tcctagttct catctcccaa 1980
agtcaagcca atgtgtgaag aaatgaccag ctcagcagcc aaggcccagg gtgcacaggt 2040
cttcgttggg agaggcatct gcaggccttt ccttgcccac tgggatcctt gcctagcata 2100
gtgacgatgt tcagccctgg agacaaacaa gaaggggaac accaacatca atagaagtat 2160
atatttacaa attgcatttc tgctgtattg aaactaacat tctgcccttt aaaatcctga 2220
aaataaaatt tcagtatgaa atgaaaaaaa                                  2250
```

<210> 52
<211> 3430
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5950519CB1

<400> 52

```
gagctgaccc tgcggggtcc cgggggggga gggggagccg cgaagccccc actgaggccg 60
ccgctgccgg gcctcccctc ccccccgggc gggcgccatg cgggggagcc cgggcgacgc 120
ggagcggcgg cagcgctggg gtcgcctgtt cgaggagctg gacagtaaca aggatggccg 180
cgtggacgtg cacgagttgc gccaggggct ggccaggctg ggcgggggca acccagaccc 240
cggcgcccaa cagggtatct cctctgaggg tgatgctgac ccagatggcg ggctcgacct 300
ggaggaattt tcccgctatc tgcaggagcg ggaacagcgt ctgctgctca tgtttcacag 360
tcttgaccgg aaccaggatg gtcacattga tgtctctgag atccaacaga gtttccgagc 420
tctgggcatt tccatctcgc tggagcaggc tgagaaaatt ttgcacagca tggaccgaga 480
cggcacaatg accattgact ggcaagaatg gcgcgaccac ttcctgttgc attcgctgga 540
aaatgtggag gacgtgctgt atttctggaa gcattccacg gtcctggaca ttggcgagtg 600
```

cctgacagtg ccggacgagt tctcaaagca agagaagctg acgggcatgt ggtggaaaca 660
gctggtggcc ggcgcagtgg caggtgccgt gtcacggaca ggcacggccc ctctggaccg 720
cctcaaggtc ttcatgcagg tccatgcctc aaagaccaac cggctgaaca tccttggggg 780
gcttcgaagc atggtccttg agggaggcat ccgctccctg tggcgcggca atggtattaa 840
tgtactcaag attgcccccg agtcagctat caagttcatg gcctatgaac agatcaagag 900
ggccatcctg gggcagcagg agacactgca tgtgcaggag cgcttcgtgg ctggctccct 960
ggctggtgcc acagcccaaa ccatcattta ccctatggag gtgctgaaga cgcggctgac 1020
cttgcgccgg acgggccagt ataaggggct gctggactgc gccaggcgta tcctggagag 1080
ggaggggccc cgtgccttct accgcggcta cctccccaac gtgctgggca tcatccccta 1140
tgcgggcatc gacctggccg tctacgagac tctgaagaac tggtggcttc agcagtacag 1200
ccacgactcg gcagacccag gcatcctcgt gctcctggcc tgcggtacca tatccagcac 1260
ctgcggccag atagccagtt acccgctggc cctggtccgg acccgcatgc aggcacaagc 1320
ctccatcgag ggtggccccc agctgtccat gctgggtctg ctacgtcaca tcctgtccca 1380
ggagggcatg cggggcctct accgggggat cgcccccaac ttcatgaagg ttattccagc 1440
tgtgagcatc tcctatgtgg tctacgagaa catgaagcag gccttggggg tcacgtccag 1500
gtgagggacc cggagcccgt ccccccaatc cctcaccccc cacacctcag ccactggaga 1560
ctgatgatcc aaccacagga tccctactct ttggccacga gatcccagta cccagatcct 1620
ggatcctaga ctcctatgcc ccaaccattg ggtcatggga tcccagcacc cagatcctgg 1680
atcctagact cctatgcccc aaccactggg tcatgcgatc cccacccttc agccactaga 1740
tcccagatcc ccctgtaacc ataactgtgg atcccttact tcagcaactc aagtctgcta 1800
ccctaaccac aagattcaag attatccaca ccccagccct taatccccat cccccaaatc 1860
actggatcct gcagccccac atcctaaggt ggatcccacg cttccctgtg cccccactg 1920
gatcctggac ctctacgtct taaccactgg atcccacaca aatcagtgaa tggatcccaa 1980
caccccaacc acaggagcac ggattccctg tacctcaaca cccagaccct gcctccctca 2040
ggcaccagat ccagtgtcct agtgaaacgc tggatcctag atccccaacc ccagatcccc 2100
atgcctcgag ccctggatct ccaagctcag ctgctggatt ctggatgtca acaaacctca 2160
ccactggatc ctgacaacca caatgcctgg atcctggggc cccatcact ggatcccaga 2220
tcccctcact ccacccactg gattcctgca ttggtttttg gtttttttgtt ttttttaac 2280
ctcgacactg ggtctcagat ccttctgctg actgccagat ccctgcattt caagcactac 2340
gccttccacc cccaggcact ggatcccaga ttcccaagcc ttcacccacc agattctggc 2400
tcctaaaaca agtgcggggg ccccagtggc acagcaagtg gatcctggca actgcagctg 2460
ctggattcca gattctgggt ccccaatccc tctgcccagt ccctcaatgt tgaaacctca 2520
tctcttgaag gcagatcctg atattccaag gcactgaatc ccaagccctg aatccccggt 2580
ttctgatctg aatcttccag gcgccgggtc ccaaatgttc aggccccaag tctagatcct 2640
ggcagcccag tcacagagta tcccacacac actggtgccc agagccggct tctcatgaca 2700
tgaaattgca tggtcgaggg agtctgtggg gaaggaagcc caggtcctgg ctgcaacctg 2760
cacggatgct ggattccccc tcaccccacc tctgcatggc caccccctcc cagccctgtg 2820
gggaaactgt tccctggaac cactccactc cctgcatccc cacacttcac agcatcttcc 2880
atccccctcc caccttctag gcgaatagtc cccagagctg tgttcctcca aggggtccga 2940
ggaatcactc actcctggag gctggcaagg agacagtctg aggccaggga cacatgaagg 3000
gatgtcccca ccccagcact atcagggcct ccccaggctt ccagagttga aagccaggag 3060
aaaatcggca aagaccaccc ttccctaaac ccaagcaccc aatgatgcaa aaaacaaaaa 3120
caaaaaaaaa ccaccaaatc cccaaattca ttccagatct atttttctac cagagagagg 3180
agcaaagtcc tcctcccctg cgcccttaca ttctgcactt catagttgga ttctgagctt 3240
aggatcatct ggagacccca tggagggact tggaaagggg aactgggatt tggggagggg 3300
ctggaggact tccgcacgct tccacctcct tcgacctcca ctgcgcccca cctccctgcc 3360
tgtgtgtgtt atttcaaagg aaaagaacaa aaggaataaa ttttctaagc tctttaaaaa 3420
aaaaaaaaaa 3430

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 March 2002 (21.03.2002)

PCT

(10) International Publication Number
WO 02/22684 A2

(51) International Patent Classification[7]: C07K 14/705

(21) International Application Number: PCT/US01/28938

(22) International Filing Date: 14 September 2001 (14.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/232,685	15 September 2000 (15.09.2000)	US
60/234,842	22 September 2000 (22.09.2000)	US
60/236,882	29 September 2000 (29.09.2000)	US
60/239,057	5 October 2000 (05.10.2000)	US
60/240,540	13 October 2000 (13.10.2000)	US
60/241,700	18 October 2000 (18.10.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **LEE, Sally** [US/US]; 825 East Evelyn #425, Sunnyvale, CA 94086 (US). **SANJANWALA, Madhu, S.** [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **RAUMANN, Brigitte, E.** [US/US]; 5801 South Dorchester Avenue #3B, Chicago, IL 60637 (US). **BRUNS, Christopher, M.** [US/US]; 575 S. Rengstorff Avenue #126, Mountain View, CA 94040 (US). **NAINI, Amir** [US/US]; 51 Spyglass Hill, Oakland, CA 94618 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **REDDY, Roopa, M.** [IN/US]; 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Pharmaceuticals, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/22684 A2

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.